Analysis of Nonlinear Behaviors, Design and Control of Sigma Delta Modulators

Charlotte Y. F. HO

Thesis submitted with the fulfillment of Master of Philosophy

Supervisors: Dr. Wolfram JUST (Department of Mathematics)
Dr. Joshua D. REISS (Department of Electronic Engineering)

Department of Mathematics & Department of Electronic Engineering
Queen Mary University of London

2013

DECLARATION

I hereby declare that I am the sole author of the thesis.

I authorize the Queen Mary College, University of London, to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize the Queen Mary College, University of London, to reproduce the thesis by photocopying or by other means, in total or in part, at the request of other institution for the purpose of scholarly research.

_____

Charlotte Yuk-Fan HO

ABSTRACT

Sigma delta modulators (SDMs) have been widely applied in analogue-to-digital (A/D) conversion for many years. SDMs are becoming more and more popular in power electronic circuits because it can be viewed and applied as oversampled A/D converters with low resolution quantizers. The basic structure of an SDM under analytical investigation consists of a loop filter and a low bit quantizer connected by a negative feedback loop.

Although there are numerous advantages of SDMs over other A/D converters, the application of SDMs is limited by the unboundedness of the system states and their nonlinear behaviors. It was found that complex dynamical behaviors exist in low bit SDMs, and for a bandpass SDM, the state space dynamics can be represented by elliptic fractal patterns confined within two trapezoidal regions. In all, there are three types of nonlinear behaviors, namely fixed point, limit cycle and chaotic behaviors. Related to the unboundedness issue, divergent behavior of system states is also a commonly discovered phenomenon. Consequently, how to design and control the SDM so that the system states are bounded and the unwanted nonlinear behaviors are avoided is a hot research topic worthy of investigated.

In our investigation, we perform analysis on such complex behaviors and determine a control strategy to maintain the boundedness of the system states and avoid the occurrence of limit cycle behavior. For the design problem, we impose constraints based on the performance of an SDM and determine an optimal design for the SDM. The results are significantly better than the existing approaches.

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATION

sigma delta modulator	SDM
analogue-to-digital	A/D
digital-to-analogue	D/A
delta modulation	DM
pulse code modulation	PCM
differential PCM	DPCM
oversampling ratio	OSR
Karush-Kuhn-Tucker	KKT
Signal-to-noise ratio	SNR
noise transfer function	NTF
signal transfer function	STF
semi-infinite programming	SIP
infinite impulse response	IIR
finite impulse response	FIR
dynamic range	DR
peak SNR	PSNR
spurious free dynamic range	SFDR
total harmonic distortion	THD
fast Fourier transform	FFT

AUTHOR PUBLICATIONS

I. Sigma Delta Modulators

Book Chapters:

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING and Joshua D. REISS, "Control of Sigma Delta Modulators via Fuzzy Impulsive Approach," *Control of Chaos in Nonlinear Circuits and Systems, Nonlinear Scientific Series of Nonlinear Science Series A*, World Scientific Publishing, vol. 64, pp. 245-270, 2009.

International Journal Papers:

Charlotte Yuk-Fan HO, Bingo Wing-Kuen Ling, Joshua D. Reiss and Xinghuo Yu, "Global stability, limit cycles and chaotic behaviors of second order interpolative sigma delta modulators," *International Journal of Bifurcation and Chaos*, vol. 21, no. 6, pp. 1755-1772, 2011.

Charlotte Yuk-Fan HO and Bingo Wing-Kuen LING, "Can a Second Order Bandpass Sigma Delta Modulator Achieve High Signal-to-noise Ratio for Lowpass Inputs," *Chaos, Solitons and Fractals*, vol. 37, pp. 928-930, 2008.

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING and Joshua D. REISS, "Difference between Irregular Chaotic Patterns of Second-Order Double-loop ΣΔ Modulators and Second-Order Interpolative Bandpass ΣΔ Modulators," *Chaos, Solitons and Fractals*, vol. 33, no. 5, pp. 1777-1782, 2007.

Charlotte Yuk-Fan HO and Bingo Wing-Kuen LING, "Stability of Sinusoidal Responses of Interpolative Sigma Delta Modulators," *Chaos, Solitons and Fractals*, vol. 32, no. 2, pp. 480-486, 2007.

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING and Joshua. D. REISS, "Estimation of an Initial Condition of Sigma-Delta Modulators via Projection Onto Convex Sets," *IEEE Transactions on Circuits and Systems-I: Regular Papers*, vol. 53, no. 12, pp.2729-2738, 2006.

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING, Joshua D. REISS and Xinghuo YU, "Nonlinear Behaviors of Bandpass Sigma Delta Modulators with Stable System Matrices," *IEEE Transactions on Circuits and Systems—II: Express Briefs*, vol. 53, no. 11, pp. 1240-1244, 2006.

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING and Joshua D. REISS, "Fuzzy Impulsive Control on Higher Order Interpolative Lowpass Sigma Delta Modulators," *IEEE Transactions on Circuits and Systems—I: Regular Papers*, vol. 53, no. 10, pp. 2224-2233, 2006.

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING, Joshua D. REISS, Yan-Qun LIU, and Kok-Lay TEO, "Design of Interpolative Sigma-Delta Modulators via Semi-infinite

Programming," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 4047-4051, 2006.

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING and Joshua D. REISS, "Stability of Sinusoidal Responses of Marginally Stable Bandpass Sigma Delta Modulators," *International Journal of Circuit Theory and Applications*, vol. 34, no. 6, pp.593-605, 2006.

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING, Joshua D. REISS and Xinghuo YU, "Occurrence of Elliptical Fractal Patterns in Multi-bit Bandpass Sigma Delta Modulators," *International Journal of Bifurcation and Chaos*, vol. 15, no. 10, pp. 3377-3380, 2005.

International Conference Papers:

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING, Joshua D. REISS and Xinghuo YU, "Noise Analysis of Modulated Quantizer based on Oversampled Signals," *Proceedings of the International Conference of Acoustics, Speech and Signal Processing*, ICASSP, vol. 3, pp. 728-731, May 2006. (Toulouse)

Bingo Wing-Kuen LING, **Charlotte Yuk-Fan HO**, Joshua D. REISS and Xinghuo YU, "Nonlinear Behaviors of Bandpass Sigma Delta Modulators with Stable System Matrices," *Proceedings of the International Conference of Acoustics, Speech and Signal Processing*, ICASSP, vol. 4, pp. 73-76, March 2005. (Philadelphia)

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING and Joshua D. REISS, "Estimation of Initial States of Sigma-Delta Modulators," *The 120th Convention of Audio Engineering Society*, AES, no. 299, May, 2006. (Paris).

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING and Joshua D. REISS, "Using SIP Techniques for the Verification of the Trade-off Between SNR and Information Capacity of the Noise Shaped Channel," *The 120th Convention of Audio Engineering Society*, AES, no. 280, May, 2006. (Paris).

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING and Joshua D. REISS, "Fuzzy Impulsive Control of Higher Order Sigma-delta Modulators," *The 118th Convention of Audio Engineering Society*, AES, no. 6451, May 2005. (Barcelona)

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING and Joshua D. REISS, "Design of Interpolative Sigma Delta Modulators via a Semi-infinite Programming Approach," *Proceeding of the 5th Advanced A/D and D/A Conversion Techniques and Their Applications*, ADDA, pp. 271-276, July 2005. (Limerick)

II. Digital Filters / Filterbanks with Nonlinearities

International Journal Papers:

Bingo Wing-Kuen LING, **Charlotte Yuk-Fan HO** and Peter Kwong-Shun TAM, "Chaotic Filter Bank for Computer Cryptography," *Chaos, Solitons and Fractals*, vol. 34, no. 3, pp. 817-824, 2007.

Bingo Wing-Kuen LING, **Charlotte Yuk-Fan HO** and Peter Kwong-Shun TAM, "Nonlinear Behaviors of First and Second Order Complex Digital Filters with Two's

Complement Arithmetic," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 4052-4055, 2006.

Bingo Wing-Kuen LING, **Charlotte Yuk-Fan HO** and Peter Kwong-Shun TAM, "Normalized Histogram of the State Variable of First-order Digital Filters with Two's Complement Arithmetic," *International Journal of Bifurcation and Chaos*, vol. 15, no. 8, pp. 2583-2586, 2005.

Bingo Wing-Kuen LING, **Charlotte Yuk-Fan HO** and Peter Kwong-Shun TAM, "Detection of Chaos in Some Local Regions of Phase Portraits Using Shannon Entropies", *International Journal of Bifurcation and Chaos*, vol. 14, no. 4, pp. 1493-1499, 2004.

Bingo Wing-Kuen LING, **Charlotte Yuk-Fan HO** and Peter Kwong-Shun TAM, "Admissibility of Unstable Second-order Digital Filter with Two's Complement Arithmetic," *International Journal of Circuit Theory and Applications*, vol. 32, no. 3, pp.97-104, 2004.

Bingo Wing-Kuen LING, **Charlotte Yuk-Fan HO**, Raymond Shing-Keung LEUNG and Peter Kwong-Shun TAM, "Oscillation and Convergence Behaviors Exhibited in an 'Unstable' Second-order Digital Filter with Saturation-type Nonlinearity," *International Journal of Circuit Theory and Applications*, vol. 32, no. 2, pp.57-64, 2004.

International Conference Papers:

Bingo Wing-Kuen LING, **Charlotte Yuk-Fan HO** and Peter Kwong-Shun TAM, "Nonlinear Behaviors of Second-order Digital Filters with Two's Complement Arithmetic," *Proceedings of the 4th ACM Postgraduate Research Day*, pp. 21-30, January 2003. (Hong Kong)

III. TCP/IP with Nonlinearities

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING, Herbert Ho-Ching IU and Tyrone L. Femando, "Nonconvex integer optimal robust early detection algorithm," *ISA transactions*, vol. 51, pp. 439-445, 2012.

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING and Herbert Ho-Ching IU, "Symbolic dynamical model of average queue size of random early detection algorithm," *International Journal of Bifurcation and Chaos*, vol. 20, no. 5, pp. 1415-1437, 2010.

International Conference Papers:

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING, Herbert Ho-Ching IU, Hak-Keung LAM and Ornifer Ekwalanga EKOZE, "Asymptotical Stability of Random Early Detection Algorithm for Internet Congestion Problem," *The 2nd International Conference on Information and Systems Sciences*, ICISS, December 2008. (Dalian)

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING and Zhi-Wei CHI, "Linear Phase FIR Two-Channel Uniform Maximally Decimated Modulated Filter Bank Design via a Weightless Multi-Criterion Functional Inequality Constrained Optimization Approach," *The 2nd International Conference on Information and Systems Sciences*, ICISS, December 2008. (Dalian)

IV. Applications of Continuous Constrained Optimization Theory

International Journal Papers:

Charlotte Yuk-Fan HO, Bingo Wing-Kuen Ling, Heidi Hai Huyen Dam and Kok-Lay Teo, "Minimax passband group delay nonlinear phase peak constrained FIR filter design without imposing desired phase response," *International Journal of Innovative Computing, Information and Control*, vol. 8, no. 5B, pp. 3863-3874, 2012.

Charlotte Yuk-Fan HO, Bingo Wing-Kuen Ling, Lamia Benmesbah, Ted Chi-Wah Kok, Wan-Chi Siu and Kok-Lay Teo, "Two-channel linear phase FIR QMF bank minimax design via global nonconvex optimization programming," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4436-4441, 2010.

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING, Yan-Qun LIU, Peter Kwong-Shun TAM and Kok-Lay TEO, "Optimum Design of Discrete-time Differentiators via Semi-infinite Programming Approach," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 2, pp. 168-172, 2008.

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING, Zhi-Wei CHI, Mohammad SHIK-BAHAEI, Yan-Qun LIU and Kok-Lay TEO, "Design of Near Allpass Strictly Stable Minimal Phase Real Valued Rational IIR Filters," *IEEE Transactions on Circuits and Systems—II: Transactions Brief*, vol. 55, no. 8, pp.781-785, 2008.

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING, Yan-Qun LIU, Peter Kwong-Shun TAM and Kok-Lay TEO, "Optimal PWM Control of Switched-Capacitor DC-DC Power Converters via Model Transformation and Enhancing Control Techniques," *IEEE Transactions on Circuits and Systems—I: Regular Papers*, vol. 55, no. 5, pp. 1382-1391, 2008.

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING, Yan-Qun LIU, Peter Kwong-Shun TAM and Kok-Lay TEO, "Efficient Algorithm for Solving Semi-Infinite Programming Problems and Their Applications to Nonuniform Filter Bank Designs," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp.4223-4232, 2006.

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING, Yan-Qun LIU, Peter Kwong-Shun TAM and Kok-Lay TEO, "Optimal Design of Magnitude Responses of Rational Infinite Impulse Response Filters," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 4039-4060, 2006.

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING, Yan-Qun LIU, Peter Kwong-Shun TAM and Kok-Lay TEO, "Optimal Design of Nonuniform FIR Transmultiplexer Using Semi-infinite Programming," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2598-2603, 2005.

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING, Yan-Qun LIU, Peter Kwong-Shun TAM and Kok-Lay TEO, "Design of Nonuniform Near Allpass Complementary FIR Filters via a Semi-infinite Programming Technique," *IEEE Transactions on Signal Processing*, vol. 53, no. 1, ptwop. 376-380, 2005.

Accepted/Published International Conference Papers:

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING, Hai Huyen DAM and Kok-Lay TEO, "Minimax passband group delay nonlinear FIR filter design without imposing desired phase response," *19th European Signal Processing Conference*, EUSIPCO, August-September 2011. (Barcelona).

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING, Zhi-Wei CHI, Chi-Wah KOK and Wan-Chi SIU, "Empirical Formula for Designing Symmetric/Anti-symmetric FIR Single Band PCLS Filters," *European Conference on Signal Processing*, EUSIPCO, 25-29 August, 2008. (Lausanne).

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING, Yan-Qun LIU, Peter Kwong-Shun TAM and Kok-Lay TEO, "Optimum Nonuniform Transmultiplexer Design," *Proceedings of the International Conference on Neural Networks and Signal Processing*, ICNNSP, pp. 740-743, December 2003. (Nanjing) (Invited Special Session)

V. Wavelet Signal Processing and Filterbanks

International Journal Papers:

Charlotte Yuk-Fan HO, Bingo Wing-Kuen Ling, Saverio Giovanni Blasi, Zhi-Wei Chi and Wan-Chi Siu, "Single step optimal block matched motion estimation with motion vectors having arbitrary pixel precisions," *American Journal of Engineering and Applied Sciences*, vol. 4, no. 4, pp. 448-460, 2011.

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING and Peter Kwong-Shun TAM, "Representations of Linear Dual Rate System via Single SISO LTI Filter, Conventional Sampler and Block Sampler," *IEEE Transactions on Circuits and Systems—II: Transactions Brief*, vol. 55, no. 2, pp. 168-172, 2008.

Charlotte Yuk-Fan HO, Tai-Chiu HSUNG, Daniel Pak-Kong LUN, Bingo Wing-Kuen LING, Peter Kwong-Shun TAM and Wan-Chi SIU, "Regularity Scalable Image Coding Based on Wavelet Singularity Detection," *International Journal of Image and Graphics*, vol. 8, no. 1, pp. 109-134, 2008.

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING, Thomas Pak-Lin WONG, Albert Yick-Po CHAN and Peter Kwong-Shun TAM, "Fuzzy Multiwavelet Denoising on an ECG Signal," *Electronics Letters*, vol. 39, no. 16, pp. 1163-1164, 2003.

International Conference Papers:

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING, Saverio Giovanni BLASI, Zhi-Wei CHI and Wan-Chi SIU, "Single step optimal block matched motion estimation with motion vectors having arbitrary pixel precisions," *International Symposium on Communication Systems, Networks and Digital Signal Processing*, CSNDSP, pp. 364-374, July 2010. (Newcastle)

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING, Lamia BENMESBAH, Ted Chi-Wah KOK, Wan-Chi SIU and Kok-Lay TEO, "Optimal cosine modulated nonuniform linear phase FIR filter bank design via stretching and shifting frequency response of prototype

filter," *International Symposium on Communication Systems, Networks and Digital Processing*, pp. 545-550, CSNDSP, July 2010. (Newcastle).

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING, Hak-Keung LAM, Thomas Pak Lin WONG, Albert Yick Po CHAN, and Peter Kwong-Shun TAM, "Fuzzy Rule Based Multiwavelet ECG Signal Processing," *International Conference on Fuzzy Systems*, IEEE-FUZZ, pp. 1064-1068, 1-6 June, 2008. (Hong Kong).

Yuk-Fan HO, Tai-Chiu HSUNG and Daniel Pak-Kong LUN, "Adaptive Wavelet Regularity Scalable Image Coding," *Proceedings of the International Conference of Acoustics, Speech and Signal Processing*, ICASSP, vol. 4, pp.3493-3496, 13-17 May 2002. (Philadelphia)

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING and Peter Kwong-Shun TAM, "Denoising by Multiwavelet Singularity Detection," *Proceedings of the International Conference on Neural Networks and Signal Processing*, ICNNSP, pp. 616-619, December 2003. (Nanjing) (Invited Special Session)

Daniel Pak-Kong LUN, Tai-Chiu HSUNG and **Yuk-Fan HO**, "Wavelet Singularity Detection for Image Processing," *Proceedings of the 45th Midwest Symposium on Circuits and Systems*, MSCAS, vol. 2, pp. 156-159, 4-7 August 2002. (U.S.A.) (Invited Lecture)

VI. Intelligent and Biomedical Systems and Perceptrons

International Journal Papers:

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING and Herbert Ho-Ching IU, "Invariant set of weight of perceptron trained by perceptron training algorithm," *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, vol. 40, no. 6, pp. 1521-1530, 2010.

Charlotte Yuk-Fan HO and Bingo Wing-Kuen LING, "Initiation of HIV therapy," *International Journal of Bifurcation and Chaos*, vol. 20, no. 4, pp. 1279-1292, 2010.

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING, Hak-Keung LAM and Muhammad H U NASIR, "Global Convergence and Limit Cycle Behavior of Weights of Perceptron," *IEEE Transactions on Neural Networks*, vol. 19, no. 6, pp. 938-947, 2008.

International Conference Papers:

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING, Muhammad Habib Ullah NASIR, Hak-Keung LAM and Herbert H. C. IU, "Characterization of Set of Vectors Represented by Lattices," *International Symposium on Communication Systems, Networks, and Digital Signal Processing*, CSNDSP, pp. 711-715, 23-25 July, 2008. (Graz).

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING, Muhammad Habib Ullah NASIR, Hak-Keung LAM and Herbert H. C. IU, "Properties of an Invariant Set of Weights of Perceptrons," *International Joint Conference on Neural Networks*, IJCNN, pp. 1630-1635, 1-6 June, 2008. (Hong Kong).

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING and Hak-Keung LAM, "Initiation and Dose Concentration of HIV Control," *The Third Shanghai International Symposium on Nonlinear Science and Applications*, NSA, pp. 160-163, June 2007. (Shanghai)

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING, Hak-Keung LAM and Muhammad H. U. NASIR, "Boundedness of Weighted Coefficients of Perceptron Learning Algorithm and Global Convergence of Fixed Point and Limit Cycle Behaviors," *The Third Shanghai International Symposium on Nonlinear Science and Applications*, NSA, pp. 58-63, June 2007. (Shanghai)

Charlotte Yuk-Fan HO, Bingo Wing-Kuen LING, Yan-Qun LIU, Peter Kwong-Shun TAM and Kok-Lay TEO, "Fuzzy Switching Systems: Minimizing Discontinuities and Ripple Magnitude and Energy," *Proceedings of the International Conference on Complex Systems, Intelligence and Modern Technology Applications*, CSIMTA, pp. 139-144, September 2004. (Cherbourg)

CHAPTER I. INTRODUCTION

Real world signals are analogue. However, analogue signals are not robust to noise because analogue signals consist of infinite number of levels. Very small amount of noise would corrupt the analogue signals. On the other hand, digital signals consist of finite number of discrete levels. If the noise level is lower than the quantization level of a digital signal, then the noise of the digital signal can be eliminated via applying the quantization on the digital signal. To obtain digital signals, continuous-time signals are first sampled into discrete-time signals, then quantization is applied on these discrete-time signals.

Digital signals are very robust to noise and easy to process, such as in storage, transmission and manipulation, because of the advanced of computer technology. However, the digitization process requires A/D conversion, while changing digital signals back to continuous-time signals require D/A conversion. Hence, A/D and D/A conversion are very important processes for many engineering applications.

SDM are widely employed in A/D conversion of audio signals. This is because human are sensitive for audio signals in the frequency band 20-20kHz. Hence, the Nyquist sampling rate is about 40kHz, and the sampling rate for audio compact disk is usually 44.1kHz. Supposing that the OSR is 64, then the sampling frequency will be 2.8224MHz, which is implementable using current inexpensive switched-capacitor circuit technology. For audio signal processing, one can generate direct stream digital audio if the input signal is a multi-bit quantized signal of the SDM [4].

Beside audio applications, SDMs are found in many low signalling rate, equivalently high resolution and relatively narrow conversion bandwidth applications, such as precision measurement devices, battery-operated communication systems, amplitude modulation communication systems and cardiac acquisition systems, etc.

Precision measurement devices in weak magnetic field measurement systems [5] use micro-fluxgate sensors that employ a sensor signal processing unit for the acquisition of data. The sensor signal processing unit involves the sigma delta modulation in the negative feedback loop. The lowpass filtered bitstream output of the SDM is then fed

back to the magnetic field system so that the system linearity, hysteresis and stability are improved.

For microwave power amplifiers [3], a bandpass SDM is employed to encode the desired output signal into a binary level signal representing an analogue radio frequency input, which is subsequently fed into a switching-mode power amplifier. A bandpass filter is used to remove unwanted spectral components from the output.

For amplitude modulation A/D conversion [6], a single-loop fourth-order bandpass SDM with a 1-bit quantizer is applied directly to A/D conversion of narrow band signals within the commercial amplitude modulation band, which is from 540kHz to 1.6MHz. The signal bandwidth is 1.06MHz. After applying the bandpass SDM, the amplitude modulation sampling frequency can be performed at 6.67MHz. The OSR can be reduced greatly to 3.1462 approximately. This demonstrates the advantages of employing bandpass SDMs in narrowband high frequency communication systems.

For the acquisition of cardiac signals in implantable pacemakers [7], the catheters connected to the cardiac muscle are AC coupled to the chip inputs by means of off-chip highpass filters. In each channel the acquired signal is amplified by a translinear front-end. Finally, bandpass filtering is performed by the translinear front-end in order to suppress noise components out of the band of interest. The signal is then converted to the digital domain by a sigma delta A/D converter. In particular, the SDM is integrated on-chip, while the corresponding decimation filter has to be designed in order to process data from measurement. The accuracy of the A/D converter strongly depends on the detection strategy used in the pacemaker.

By modeling the quantizer as an additive white noise source, the SDM can be modeled by a two-input one-output linear time-invariant system. The block diagram of the basic structure of a SDM is shown in Figure 1.1 and the dynamics of the SDM is governed by the following dynamical equation:

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{B}(\mathbf{u}(k) - Q(\mathbf{x}(k))),$$

where $\mathbf{x}(k)$ and $\mathbf{u}(k)$ are the state vector and the input, respectively, $\mathbf{A}$ and $\mathbf{B}$ are constant matrices, and $Q(\cdot)$ is a single bit quantization function. The loop filter can be then designed so that a small value of the noise transfer function (NTF) and an approximate unity gain of the signal transfer function (STF) are achieved at the signal

band. This will result to a minimum overlapping between the noise spectrum and the input spectrum. Consequently, the SDM can achieve a very high signal-to-noise ratio (SNR).

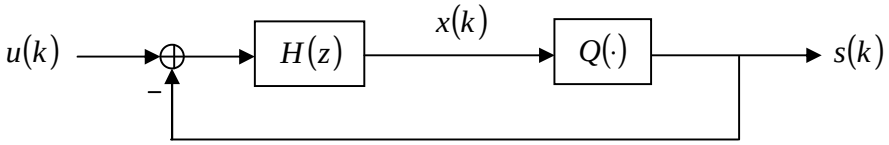


Figure 1.1 The block diagram of a SDM.

In this chapter, we will first give an historical overview of the development of SDM. Then we will study the background on SDMs and their nonlinear behavior in Section 1.2 and Section 1.3 respectively. Next, we will present a literature review on the existing research works and explain briefly how they are not sufficient to address the existing research problems in Section 1.4. Finally, we will give an overview of the thesis in Section 1.5.

1.1 Historical Overview

The research of SDM is originated from Delta modulation (DM) [9], [10]. DM is a kind of source coding technique and its block diagram is shown in Figure 1.2. This technique was devised in the 1940s as an alternative to pulse code modulation (PCM) coding. Though DM has been known for a long time, it gained interest during the 1980s due to its utilization in the design of PCM codec.

DM is a single bit version of differential PCM (DPCM) in which the error or the difference between the input signal and the output signal is quantized by a one-bit quantizer and represented by rectangular pulses. The output signal is the integration of these rectangular pulses. In this sense, the output signal approximates the input signal. Since the output is an integral of rectangular pulses, it is a piecewise ramp type signal and it does not consist of any discontinuity.

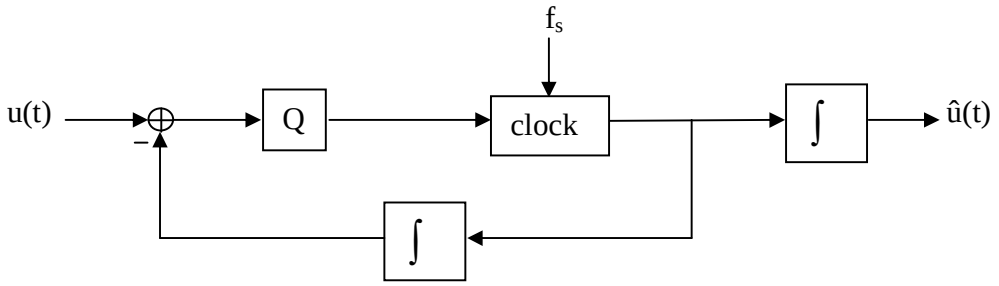


Figure 1.2 The block diagram of a delta modulator.

Sigma delta modulation is also a kind of source coding technique. The name "delta sigma" was used by some of the earliest researchers in the field, such as Inose and Yasuda in 1963 [12], but now the term "sigma delta" has become almost synonymous with noise shaping analogue-to-digital (A/D) conversion, as discussed by Aziz *et. al.* in 1996 [2]. The word "delta" in "delta sigma" refers to the narrow shape of the input spectrum, which is approximately as a delta function, after oversampling.

Sigma delta modulation has been proposed and applied for more than 40 years and it was first patented by Cutler in 1960 [11]. This technique is further elaborated by Brahm in 1965 [13] and Miura *et al.* in 1971 [14]. It was then proposed to be applied in A/D converter by Candy in 1974 [1].

1.2 Background on SDMs

In this section, we will introduce two important concepts of the working principle of SDM in Section 1.2.1 and 1.2.2 after introducing the SDM in Section 1.2.1. They are the oversampling principle and the noise shaping principle. For the oversampling principle, we will show that when the oversampling ratio (OSR) increases, the signal-to-noise (SNR) will increase.

1.2.1 The structure of SDMs

SDM is a closed loop system consisting of a linear time-invariant loop filter and a memoryless quantizer. The components of feedforward SDMs are connected by a negative feedback loop. The structure is very simple and it is usually implemented via switched-capacitor circuits. It consists of two operations: filtering the signals in frequency domain and quantizing the signals in magnitude.

There are two basic types of SDMs: the interpolative or feedforward and the feedbackward structures. In this research, only the feedforward structure is considered. That is, there is only one feedback signal inputted to the loop filter as shown in Figure 1.1. The order of an SDM is equal to the order of its loop filter. The number of bits of the SDM is equal to the number of bits of the quantizer.

1.2.2 The oversampling principle

It is well known that a continuous-time signal can be sampled into a discrete-time signal and perfect reconstruction can be achieved via an ideal filtering if the sampling rate is higher than twice the bandwidth of the corresponding continuous-time signal. This sampling rate is called the Nyquist sampling rate [28]. When the sampling rate is much higher than the Nyquist sampling rate, the sampled signal is referred as an oversampled signal and the process is known as oversampling. The ratio of the oversampled frequency to the Nyquist sampling rate is called the oversampling ratio (OSR). The principle of oversampling in A/D conversion can be found in the work by Hauser in 1991 [29]. After oversampling the continuous-time input signal, the spectrum of the sampled signal is a periodic version of the original input spectrum, where the period in the frequency spectrum is directly proportional to the product of the OSR and the Nyquist sampling rate. By converting the sampled signal to discrete-time sequences, the spectrum of the discrete-time sequences is mapped to a 2π-periodic spectrum. Hence, the spectrum of the discrete-time sequences will become narrower as the OSR increases. Consequently, the noise effect from the quantizer becomes less significant.

To understand this phenomenon, we assume that the quantization noise is uncorrelated to the input signal of the quantizer, so the power spectral density of the quantization noise is uniformly spread over the whole spectrum $(-\pi, \pi)$. This uniform spread occurs because the correlation function of any two uncorrelated signals is a discrete-time delta function and the discrete-time Fourier transform of a delta function is constant. Denote the noise magnitude in the power spectral density as N_0. Denote the Nyquist sampling rate, OSR, continuous-time input signal and continuous-time sampled

signal as, respectively, f_s, R, $x(t)$ and $x_s(t)$. Then the sampling period is $\frac{1}{Rf_s}$ and we have

$$x_s(t) = x(t)\sum_{n=-\infty}^{+\infty}\delta\left(t-\frac{n}{Rf_s}\right).$$

Denote the continuous-time Fourier transform of $x_s(t)$ as $X_s(\omega)$. As the continuous-time Fourier transform of $\sum_{n=-\infty}^{+\infty}\delta(t-nT)$ is $\frac{2\pi}{T}\sum_{n=-\infty}^{+\infty}\delta\left(\omega-\frac{2\pi n}{T}\right)$ where $T=\frac{1}{Rf_s}$, we have the continuous-time Fourier transform of $\sum_{n=-\infty}^{+\infty}\delta\left(t-\frac{n}{Rf_s}\right)$ being equal to $2\pi Rf_s\sum_{n=-\infty}^{+\infty}\delta(\omega-2\pi Rf_s n)$. By applying the convolution property of the continuous-time Fourier transform, that is $X_s(\omega)=\frac{1}{2\pi}X(\omega)*2\pi Rf_s\sum_{n=-\infty}^{+\infty}\delta(\omega-2\pi Rf_s n)$, where $X(\omega)$ is the continuous-time Fourier transform of $x(t)$, we have

$$X_s(\omega) = Rf_s\sum_{n=-\infty}^{+\infty}X(\omega-2\pi Rf_s n).$$

Since

$$x_s(t) = x(t)\sum_{n=-\infty}^{+\infty}\delta\left(t-\frac{n}{Rf_s}\right) = \sum_{n=-\infty}^{+\infty}x\left(\frac{n}{Rf_s}\right)\delta\left(t-\frac{n}{Rf_s}\right),$$

by taking the continuous-time Fourier transform on $x_s(t)$, we have

$$X_s(\omega) = Rf_s\sum_{n=-\infty}^{+\infty}X(\omega-2\pi Rf_s n) = \sum_{n=-\infty}^{+\infty}x\left(\frac{n}{Rf_s}\right)e^{-\frac{j\omega n}{Rf_s}}.$$

As the discrete-time sequences are $x\left(\frac{n}{Rf_s}\right)$, taking the discrete-time Fourier transform on this discrete-time sequences, which are denoted as $X_D(\omega)$, we have

$$X_D(\omega) = \sum_{n=-\infty}^{+\infty}x\left(\frac{n}{Rf_s}\right)e^{-j\omega n}.$$

Hence, we have

$$X_D\left(\frac{\omega}{Rf_s}\right) = X_s(\omega) = Rf_s \sum_{n=-\infty}^{+\infty} X(\omega - 2\pi Rf_s n).$$

Since the Nyquist sampling rate is f_s, $X(\omega)$ is bandlimited within $(-\pi f_s, \pi f_s)$. As a result, $X_D(\omega)$ is bandlimited within $\left(-\frac{\pi}{R}, \frac{\pi}{R}\right)$. Consequently, the noise power corrupted to the input signal is $\frac{2\pi N_0}{R}$, which is inversely proportional to R. Hence, the signal-to-noise ratio (SNR) is directly proportional to R. As a result, the plot of the OSR against the SNR will correspond to a straight line as shown in Figure 1.3. Note that it is assumed that there is no correlation between the quantization noise and the input of the quantizer.

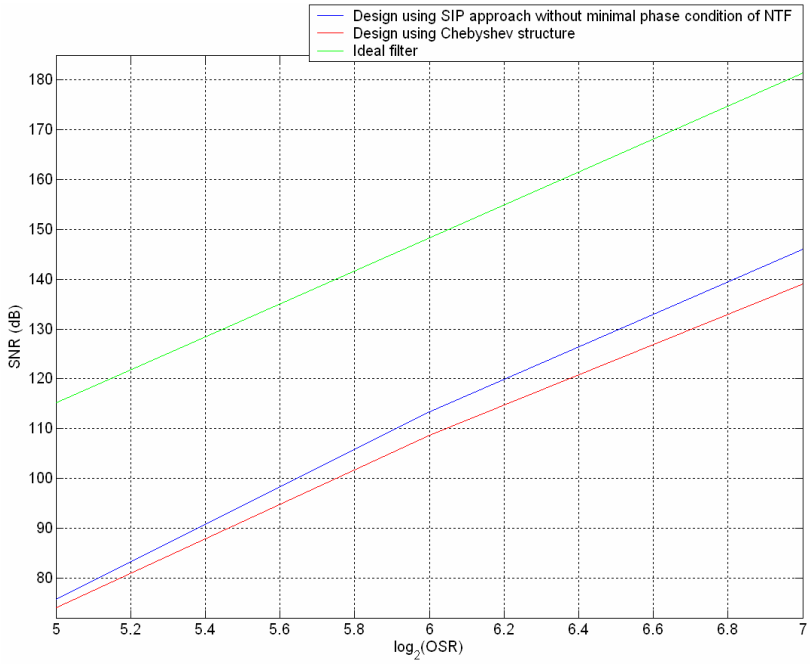


Figure 1.3 The relationship between SNR and OSR of a SDM.

By employing the oversampling technique, the correlation between the adjacent samples can be greatly increased and the change between two adjacent samples becomes very small and insignificant. Consequently, tracking of the input signal performed via the negative feedback loop that subtracts the difference between the input signal and the output signal from the quantizer can be performed accurately. This technique is very useful for further filtering and decimation processes and widely adopted in sigma delta modulation.

1.2.3 The noise shaping principle and the high SNR performance

By modeling the quantizer as an additive white noise source, the SDM can be modeled by a two-input one-output linear time-invariant system as shown in Figure 1.4, and the noise transfer function (NTF) and the signal transfer function (STF) can be defined accordingly. The loop filter can be designed so that a small value of the NTF and an approximate unity gain of the STF are achieved at the signal band. This will result to a minimum overlapping between the noise spectrum and the input spectrum. Consequently, the SDM can achieve a very high SNR. The noise is thus shaped away from the input signal and this technique is known as noise shaping technique [11].

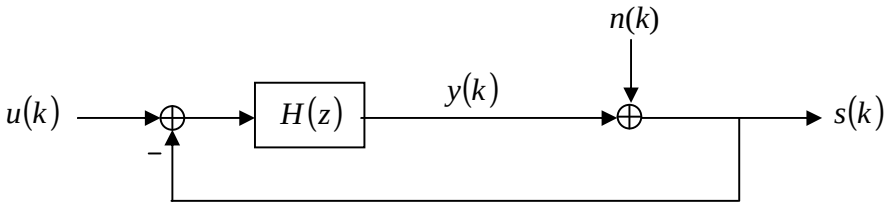


Figure 1.4 The basic structure of a SDM.

To understand the noise shaping principle, denote the z-transform of the input of the SDM $u(k)$, the output of the loop filter $y(k)$, the output of the SDM $s(k)$ and the quantization noise $n(k)$ as $U(z)$, $Y(z)$, $S(z)$ and $N(z)$, respectively. Hence, we have

$$H(z)(U(z)-S(z))=Y(z)$$

and

$$S(z)=Y(z)+N(z).$$

This implies that

$$S(z)=H(z)(U(z)-S(z))+N(z).$$

In other words,

$$S(z)=H(z)U(z)-H(z)S(z)+N(z)$$

$$S(z)=\frac{H(z)U(z)}{1+H(z)}+\frac{N(z)}{1+H(z)}$$

Hence, the NTF and STF are, respectively,

$$\text{NTF}=\frac{S(z)}{N(z)}=\frac{1}{1+H(z)}$$

and

$$\text{STF} = \frac{S(z)}{U(z)} = \frac{H(z)}{1+H(z)}.$$

Because of the oversampling and noise shaping techniques, a SDM can achieve a very high SNR even for a very coarse quantizer, in which the SNR is calculated as the following [74]: a sine wave with frequency f_0 and amplitude A is taken as the input signal, where f_0 is located at $\frac{2}{3}$ of the bandwidth of the input signal. SNR is defined as the ratio of the energy of the output of SDM at frequency f_0 to that of the sum of other frequencies within the passband of the loop filter. That is,

$$\text{SNR} = \frac{\frac{A^2}{2}}{\int_{-\frac{\pi}{R}}^{\frac{\pi}{R}} |S(\omega)|^2 d\omega},$$

where $S(\omega)$ is the frequency spectrum of the output sequences. As a result, SDMs are employed as A/D converters in many circuits and systems. For example, the magnitude of NTF at the signal band of a fifth order infinite impulse response (IIR) loop filter can be as low as 10^{-6} or -120dB.

In this research work, various performance measures are evaluated and some of these performance measures are used as the criteria in the optimal design. The most common performance measures are SNR, dynamic range (DR) and peak SNR (PSNR). Other performance indices, such as the ratio of the total power of the shaped quantization noise to that of the unshaped quantization noise, the measure of the total loss of channel information capacity after noise shaping, the noise shaping characteristics, the relationships between the SNR and the OSR, that of the number of bits of the quantizer and the filter order, as well as the spurious free dynamic range (SFDR) and the total harmonic distortion (THD) can also be employed to evaluate the three design approaches: the SIP approach, the Chebyshev structures and the Butterworth structures.

The metric that we usually employed for comparing the performances of different A/D converters is SNR. The theoretical limit of SNR of an SDM may be estimated from the work in [2],

$$\text{SNR}_{\text{estimated}} = 10\log_{10}\left(\sigma_u^2\right) - 10\log_{10}\left(\sigma_n^2\right) - 10\log_{10}\left(\frac{\pi^{2N}}{2N+1}\right) + (20N+10)\log_{10} R \text{ (dB)}, \quad (1.1)$$

where σ_u^2 and σ_n^2 are the variance of the input signal and the quantization noise, respectively, and N and L are the filter order and the number of bits of the quantizer, respectively. By assuming that the quantization noise is uncorrelated to the input of the quantizer, the spectrum of the quantization noise becomes flat. That is, $\sigma_n^2 = \int_0^{\Delta}\left(x - \frac{\Delta}{2}\right)^2 \frac{1}{\Delta} dx = \frac{\Delta^2}{12}$, where Δ is the quantization step size. If the saturation level is 1, then the unsaturation range of the quantizer is 2. As there are $2^L - 1$ quantization levels, $\Delta = \frac{2}{2^L - 1}$ and $\sigma_n^2 = \frac{1}{3\left(2^L - 1\right)^2}$. This implies that

$$\begin{aligned}
\text{SNR}_{\text{estimated}} &= 10\log_{10}\left(\sigma_u^2\right) - 10\log_{10}\left(\frac{1}{3\left(2^L-1\right)^2}\right) - 10\log_{10}\left(\frac{\pi^{2N}}{2N+1}\right) + (20N+10)\log_{10} R \\
&= 10\log_{10}\left(\sigma_u^2\right) + 10\log_{10}\left(3\left(2^L-1\right)^2\right) - 10\log_{10}\left(\frac{\pi^{2N}}{2N+1}\right) + (20N+10)\log_{10} R \\
&= 10\log_{10}\left(\sigma_u^2\right) + 10\log_{10}(3) + 20\log_{10}\left(2^L-1\right) - 10\log_{10}\left(\frac{\pi^{2N}}{2N+1}\right) + (20N+10)\log_{10} R \\
&= 10\log_{10}\left(\sigma_u^2\right) + 20\log_{10}\left(2^L-1\right) - 10\log_{10}\left(\frac{\pi^{2N}}{2N+1}\right) + (20N+10)\log_{10} R + 4.77
\end{aligned}$$

From the above, we can see that the SNR can be increased by increasing the number of bits of the quantizer. Besides, the SNR of an SDM can be improved by increasing the OSR. For the effect of the loop filter, it is shown that the higher the filter order, the better the noise shaping characteristics can be performed. Hence, the SNR can also be improved when the filter order is increased.

In comparison to the PCM, where there is no feedback, the SNR [2] does not relate to the filter order, and

$$\text{SNR} = 10\log_{10}\left(\frac{\sigma_u^2}{\sigma_n^2}\right). \quad (1.2)$$

As $\Delta = \frac{2}{2^L - 1}$ and $\sigma_n^2 = \frac{1}{3(2^L - 1)^2}$, so

$$\begin{aligned}\text{SNR} &= 10\log_{10}\left(\frac{\sigma_u^2}{\sigma_n^2}\right) \\ &= 10\log_{10}\left(\sigma_u^2\right) - 10\log_{10}\left(\frac{1}{3}\left(\frac{1}{2^L - 1}\right)^2\right) \\ &= 10\log_{10}\left(\sigma_u^2\right) - 10\log_{10}\left(\frac{1}{3}\right) + 10\log_{10}\left(\left(2^L - 1\right)^2\right)\end{aligned}$$

When L is large, then

$$\begin{aligned}\text{SNR} &\approx 10\log_{10}\left(\sigma_u^2\right) - 10\log_{10}\left(\frac{1}{3}\right) + 10\log_{10}\left(2^{2L}\right) \\ &= 10\log_{10}\left(\sigma_u^2\right) + 4.77 + 6.02L\end{aligned}.$$

Compared to that in SDM (equation 1.1), SDM could achieve better SNR because the noise shaping technique could further improve the SNR.

1.3 Background on nonlinear behaviors of SDMs

1.3.1 Reasons for the nonlinear analysis of SDMs

The input-output characteristic of an SDM is nonlinear and in most practical case the analogue input signal is not known a priori. To tackle this problem, most of the existing analysis usually made some assumptions so that the analysis is tractable [18], [31].

The most common assumptions are follows. First of all, the quantization error is a random signal, and the quantizer can be modeled by an additive white noise source [32]. The error sequence is a sample sequence of a wide-sense stationary white noise process, with each sample being uniformly distributed over the range of the quantization error. Secondly, The error sequence is uncorrelated with its corresponding input sequence. Thirdly the input sequence is a sample sequence of stationary random process. Forthly, the analogue input samples are within the full-scale range. Hence, there is no saturation error at the converter output. Therefore, traditional methods for the analysis of the SDM only work under the above assumptions. However, these assumptions hold for some input signals, but fail in many situations.

It was reported in [25] that elliptic fractal patterns may be exhibited on the phase plane of a bandpass SDM. In this research, we explain this phenomenon. At the same time, the set of initial conditions that generates fractal behaviors is characterized. The analysis is important because we cannot employ SDMs confidently in real applications before we perform the nonlinear analysis on the SDMs.

Moreover, some nonlinear behaviors, such as limit cycle behavior, would affect the normal operation of an SDM and cause degradation to its performance. This problem is particularly serious in audio applications. This is because limit cycle behavior corresponds to periodic output sequences that generate annoying audible tones. Hence, it is important to characterize these nonlinear behaviors so that we are able to design and operate the SDMs properly with better performances as well.

Furthermore, system states of SDMs may be bounded for some initial conditions, while unbounded for other initial conditions. Even though the STF and NTF are stable, the system states could be unbounded. Besides, the magnitude of the output of the loop filter is bounded at low input magnitude, while overloading the input magnitude would cause an increase in quantization error. Further increase of the input magnitude would cause unbounded state behaviors. This leads to a sudden drop in SNR. For the case when the system states are unbounded, the SDM will be damaged and serious hazards may occur. Hence, it is important to characterize the conditions for the occurrence of bounded system states so that serious hazards would never occur. However, the above phenomena are not found in linear systems. This is because the boundedness property of the system states of linear systems does not depend on the input signals and the initial conditions of the systems, in which it only depends on the closed loop poles of the system. It only depends on the filter coefficients. If NTF and STF are stable, then the linear system theory predict that the system states will be bounded for all initial conditions and larger but bounded inputs.

When the above nonlinear phenomena occur, that is, the phase plane exhibiting elliptic fractal patterns, the limit cycle behavior occurring or the system states being unbounded for stable STF and NTF, the output signal will no longer be random and all statistical assumptions are failed because the quantization error is no longer uncorrelated with the input signal. The correlation statistics is very complex and it is difficult to be

expressed analytically. Consequently, the linear statistical model fails to explain these nonlinear phenomena. Indeed, nonlinear techniques should be employed to analyze these nonlinear phenomena. In particular, nonlinear techniques dealing with the relationship between the nonlinear behaviors and the input signal, initial condition, loop filter coefficients as well as the structure of the SDM should be employed for the analysis.

In addition, linear approach sometimes causes problems in the design of SDMs. For example, consider the design of a second order interpolative bandpass SDM with the peak frequencies located at $\pm\frac{\pi}{3}$. As the peak frequencies are located at $\pm\frac{\pi}{3}$ and $z=e^{j\omega}$, this implies that the poles of the loop filter are $e^{\frac{j\pi}{3}}$ and $e^{-\frac{j\pi}{3}}$. Hence, the denominator of the loop filter transfer function is $\left(1-e^{\frac{j\pi}{3}}z^{-1}\right)\left(1-e^{-\frac{j\pi}{3}}z^{-1}\right)$. Assume that the OSR of the SDM is 64 and a one-bit quantizer is employed. Then by formulating the loop filter numerator coefficient design problem as an optimization problem with cost function minimizing the energy of the NTF, and solving the corresponding optimization problem via the Matlab optimization toolbox, we obtain the numerator coefficients of the loop filter and the loop filter transfer function is

$$H(z)=\frac{10^7\left(0.11z^{-1}-1.23z^{-2}\right)}{\left(1-e^{\frac{j\pi}{3}}z^{-1}\right)\left(1-e^{-\frac{j\pi}{3}}z^{-1}\right)}.$$

By putting $z=e^{j\omega}$ into the transfer function, we obtain Figure 1.5 which shows the magnitude response of the loop filter. It can be seen from the figure that the peak frequencies are located at $\pm\frac{\pi}{3}$. By using the delays of the output of the loop filter as the state variables, that is, $\mathbf{x}(k)\equiv\left[y(k-2)\quad y(k-1)\right]^T$, in which $\mathbf{x}(k)$ and $y(k)$ represent the system state vector and the output of the loop filter, respectively, and the bandpass SDM is assumed to be initially at rest, that is $\mathbf{x}(0)=\mathbf{0}$, it can be checked easily that the bandpass SDM can achieve a very high SNR. However, note that the numerator coefficients of the loop filter will become too large for the implementation. This is because the criterion for linear design approach is to achieve small value of NTF at the

signal band. Since $\mathrm{NTF} = \dfrac{1}{1+H(z)}$, if $\mathrm{NTF} \to 0$, then $|H(z)| \to +\infty$. Hence, the linear design approach would result to a set of very large numerator coefficients.

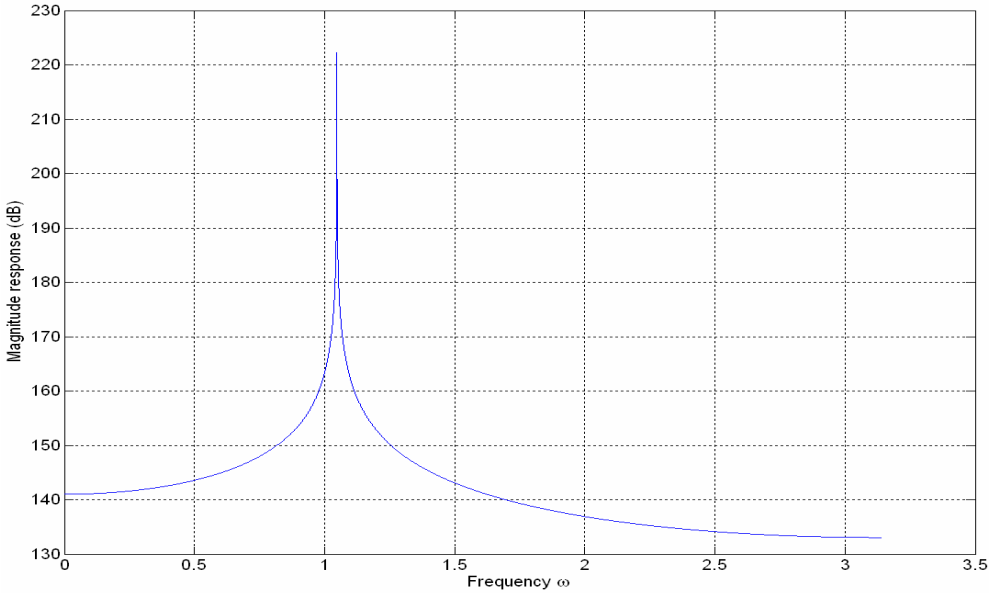


Figure 1.5 Magnitude response of the loop filter of a second order interpolative bandpass SDM.

One further example that illustrates the existence of problems in employing the linear design approach is the ideal filter. An ideal filter can achieve perfect noise shaping characteristics. However, it can cause the system states of the SDM to be unbounded that leads to a very low SNR. This shows that in order to design SDMs which can achieve high SNRs with realizable filter coefficients, investigating the noise shaping characteristics is not sufficient. It is revealed that the conditions for exhibiting global boundedness of the system states and the conditions for exhibiting various nonlinear behaviors are required to be investigated.

Note that the analysis of these nonlinear phenomena is very challenging. This is because the input-output function of the quantizer is discontinuous with the discontinuity located at the origin. This results to the occurrence of fractal and chaotic behaviors. However, many existing theorems, such as Lyapunov stability theorem, cannot be applied to explain these behaviors. This is because Lyapunov stability theorem requires the rate of the change of the Lyapunov candidate function to be monotonic decreasing and this usually results to asymptotical convergence of the state trajectory to the equilibrium point.

1.3.2 Sensitivity of the initial condition to the boundedness of system states

In this research, the local boundedness property of the system states of an SDM refers to that for a certain set of initial conditions at given input signal and filter coefficients, while the global boundedness property of the system states of an SDM refers to that for all initial conditions in the whole state space at given input signal and filter coefficients. In this research, both the local and global boundedness properties of the system states of an SDM will be analyzed. Although the output of the SDM is always bounded because of the quantizer, the boundedness of the system states of the SDM is not guaranteed and it depends on the initial condition.

As discussed in Section 1.2.1, most SDMs are implemented via switched-capacitor circuits. A slight change in the charges stored in any of the capacitors, such as leakage, would cause a corresponding change in the voltage across the capacitor. This is because the voltage across a capacitor is equal to the charge stored in it divided by its capacitance ($V = \frac{Q}{C}$). A change of the initial condition may result to an undesirable response of a nonlinear system. For example, consider the dynamics of the SDM governed by the following dynamical equation:

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{B}(\mathbf{u}(k) - Q(\mathbf{x}(k))),$$

where $\mathbf{x}(k)$ and $\mathbf{u}(k)$ are the state vector and the input, respectively, **A** and **B** are constant matrices, and $Q(\cdot)$ is a single bit quantization function. There is a change of the charges stored in the switched-capacitor corresponding to the first state variable, so that the initial condition of the first state variable now becomes 10^{-3}, that is, $\mathbf{x}(0) = \begin{bmatrix} 10^{-3} & 0 & 0 & 0 & 0 \end{bmatrix}^T$. Figure 1.6a and 1.6b show the responses of the SDMs based on the above dynamical equation with $\mathbf{u}(k) = 0.5\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \end{bmatrix}^T$ for $k \geq 0$ when $\mathbf{x}(0) = \mathbf{0}$, and $\mathbf{x}(0) = \begin{bmatrix} 10^{-3} & 0 & 0 & 0 & 0 \end{bmatrix}^T$, respectively. We can see that the state variable of the later SDM diverges, while that of the former SDM is still bounded. This example illustrates that the initial condition does affect the boundedness of the SDM.

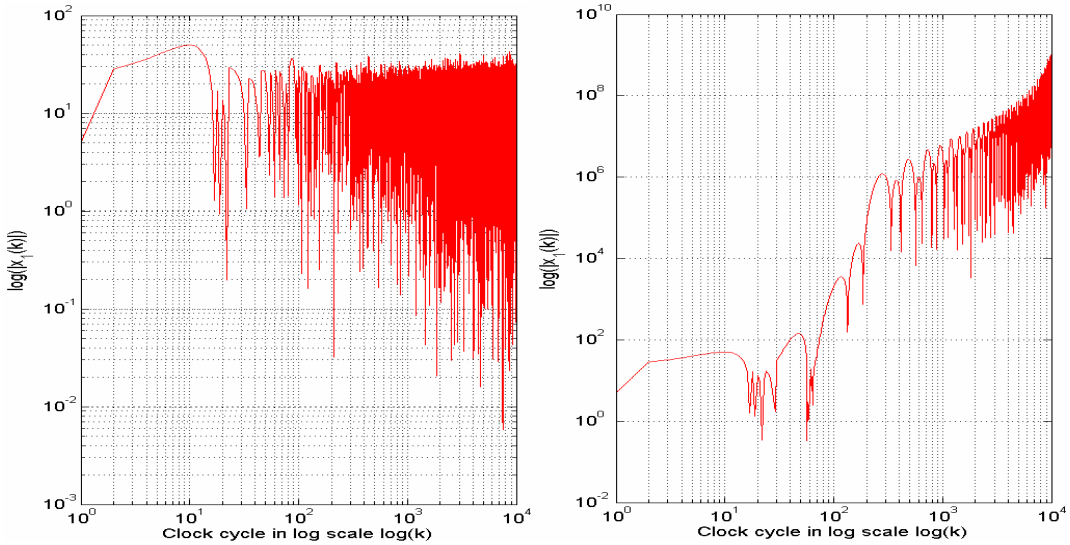


Figure 1.6 Response of the first state variable $x_1(k)$ of a fifth order lowpass SDM with (a) zero initial condition for all state variables; (b) initial condition of the first state variable being 10^{-3} and that on the remaining state variables being zero.

Some people consider that only the local boundedness of the system states is important, and the global boundedness of the system states can be ignored. However, from the example shown above, it reveals that the local boundedness of the system states is not sufficient because only a small change of the initial condition could result to the failure of the boundedness of the SDM.

1.3.3 Reasons for studying bandpass SDMs with DC input

In this research, both lowpass and bandpass SDMs are considered. We study bandpass SDMs because many systems, such as OFDM systems [67], AM and FM radio systems [68], etc, are required to perform A/D conversion on bandpass signals. By using bandpass SDMs, simple and relatively low precision analog components could achieve the objectives. Because of this advantage, this area draws much attention from the researchers in the community, and various methods for the analysis [34] and designs of bandpass SDMs [51],[52],[67],[68],[69],[70],[71] have been proposed.

In this research, we assume that the loop filter is rational, real, proper and causal, as well as there is a delay element multiplying in the numerator of the transfer function. We make these assumptions due to some implementation reasons and the feedback loop

configuration. Hence, the transfer function of a second order loop filter satisfying the above property can be represented as

$$F(z) \equiv \frac{Gz^{-1}\left(1-bz^{-1}\right)}{\left(1-a'z^{-1}\right)\left(1-az^{-1}\right)},$$

where a and a' are the poles, b is the zero and G relates to the DC gain of the loop filter. Now, let us consider the second order bandpass SDM which is discussed in [25]. In this case, $G = 2\cos\theta$, $a = e^{j\theta}$, $a' = e^{-j\theta}$ and $b = \frac{1}{2\cos\theta}$, in which θ is the natural frequency of the loop filter. Although it is a bandpass filter, the magnitude response of this bandpass filter is close to that of the lowpass filter when θ is close to zero. This is because the magnitude response of this second order bandpass filter is monotonic increasing for $\omega \in (0, \theta)$, while it is decreasing for $\omega \in (\theta, \pi)$. Hence, the frequency band $[-\theta, \theta]$ can be regarded as the passband of the filter. It is worth noting that the DC gain of this bandpass filter is not necessarily equal to zero. In fact, the DC gain of this bandpass filter is $\frac{2\cos\theta - 1}{2(1-\cos\theta)}$, while that of the lowpass filter is infinity. If $|\theta| < \cos^{-1}\frac{2\overline{H}+1}{2(1+\overline{H})}$, where $\overline{H}$ is the desired DC gain, then the bandpass filter will amplify a signal in the passband with the gain greater than $\overline{H}$. Besides, the magnitude response of this bandpass filter is very close to that of the lowpass filter when $|\omega| \geq 2\theta$. Hence, the approximation of the lowpass filter by this bandpass filter is valid when the natural frequency of this bandpass filter is close to zero. Figure 1.7 shows the magnitude responses of a second order bandpass filter with $\theta = 0.001$ and a second order lowpass filter with $a = a' = 1$, $b = \frac{1}{2}$ and $G = 2$. It can be seen from Figure 1.7 that the magnitude responses of these two filters are almost the same when $|\omega| \geq 2\theta$.

There are some reasons for us to deal with constant or DC input for the nonlinear analysis if the natural frequency of the bandpass SDM is close to zero (Figure 1.7). According to the noise shaping principle, the frequency component of the input signal should be around the peak frequency of the filter, in which the peak frequency is the frequency where the magnitude response of the filter is the largest. As the peak frequency

of the filter is located at its natural frequency, which is close to zero, dealing with DC input is appropriate.

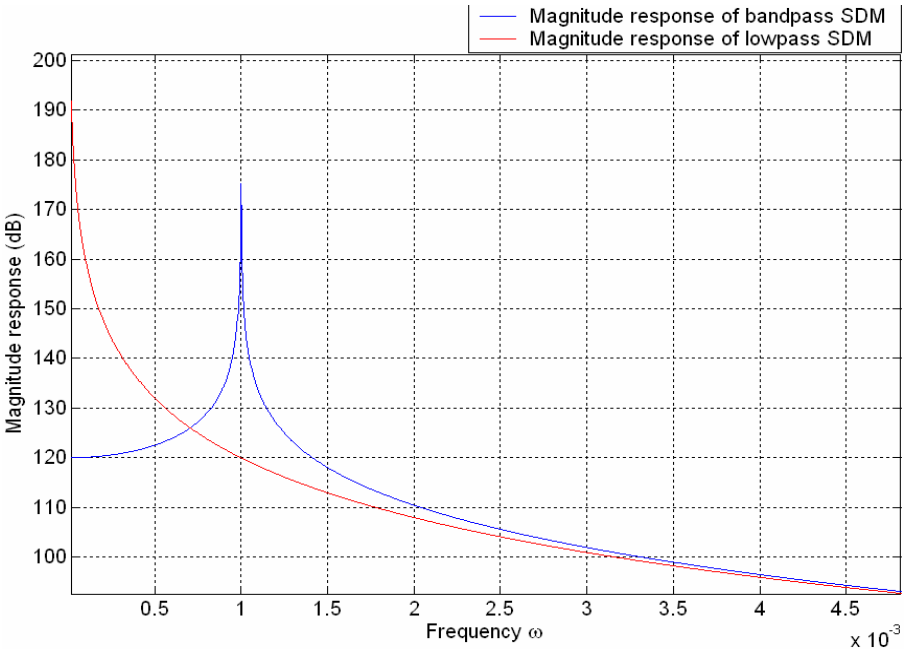


Figure 1.7 Magnitude responses of both a second order bandpass and lowpass SDMs.

One advantage of employing this bandpass SDM over the corresponding lowpass SDM is that the global boundedness of the system states can be guaranteed and the analytical proof can be found in [76]. Moreover, the conditions for this bandpass SDM exhibiting various nonlinear behaviors can be characterized and we will show the analysis in the coming chapters. Hence, by utilizing these conditions, higher SNR may be achieved.

1.4 Literature Review

Many well-known and early works have been performed for the analysis of the nonlinear behaviors exhibited in SDMs. In 1966, Smith had investigated the distortion component [17] based on modelling the quantization error as a noise source. This approach was also reviewed and discussed by Ardalan *et. al.* in 1987 [18]. However, there are some disadvantages for this method, as discussed in Section 1.3.1.

In 1952, Booton had proposed a quasi-linear method [15] for modelling nonlinear dynamical systems. Booton's quasi-linear method is based on the describing function approach [16], which assumes that the input and output signals of the quantizer are

sinusoidal signals. This method was also employed for the analysis of SDMs in [20]. The maximum input step size that a second order lowpass SDM produces limit cycle behavior was found by Hein and Zahkor in 1993 [20]. They extended this result to ascertain the existence of limit cycles in higher order lowpass SDMs via the describing function approach [19]. However, there are some disadvantages of employing these methods for analyzing nonlinear behaviors of SDMs. This method takes two assumptions: the input and output signals of the quantizer are sinusoidal, and the frequency of the input signal is the same as that of the output of the quantizer. Nevertheless, these two assumptions are not always true [21] even when a limit cycle exists.

It was found by Friedman in 1988 [30] that if the input step size of a lowpass SDM is a rational fraction of the quantizer step, then a limit cycle will occur and the period of the limit cycle will be a multiple of the denominator of that rational number.

Hyun *et. al.* proposed an algorithm in 2002 [39] for searching admissible periodic output sequences for lowpass SDMs. However, the computational complexity will certainly be increased if the order of the SDM is increased. Research is still continuing on the determination of the maximum value of the input step size that does not give rise to limit cycle behavior. This information is important for the SDM designers to avoid the occurrence of limit cycle behavior in both the lowpass and bandpass SDMs.

In 1997, Feely explored the nonlinear behavior of a bandpass SDM. The state trajectories were simulated [25] and it was found that elliptic fractal patterns would be exhibited in the phase plane. Computer simulation was performed in [25] and we analyze this behavior in this research in Section 3.3.

In 1988, Chua investigated the nonlinear behaviors of digital filters with two's complement arithmetic via symbolic dynamics [26]. Here, the two's complement arithmetic refers to the arithmetic that the most significant bit of the number is the sign bit, while the other bits represent the magnitudes of the number. Addition of two positive numbers may become negative and sum of two negative numbers may become positive. This phenomenon is called the overflow. He showed that nonlinear behaviors, such as limit cycle and elliptic fractal behaviors, would occur when overflow occurs. Since the system is a nonlinear system, the dynamics of the system can be represented by a symbolic dynamical equation. Feely used similar form of symbolic dynamical equation to

represent the dynamics of the SDMs. Hence, the behaviors occurred in digital filters with two's complement arithmetic would also occur in SDMs.

In 2003, Ling also applied symbolic dynamical approach to further analyze the necessary and sufficient conditions among the nonlinear dynamical behaviors of the digital filter with two's complement arithmetic, the periodic properties of the symbolic sequences and the corresponding sets of initial conditions [27]. When the period of the symbolic sequence is one, the state trajectory will exhibit a single ellipse. When the period of the symbolic sequence is larger than one, more than one ellipses will be exhibited on the phase portrait. These two cases are regarded as limit cycle behaviors. When the symbolic sequence is aperiodic, elliptic fractal pattern will be exhibited on the phase portrait. It is found that this theory can also be applied to explain Feely's observation and some new interesting results are discovered in this research.

However, there are some fundamental differences between the dynamics of digital filters with two's complement arithmetic and those of SDMs. For examples, the dynamics of digital filters with two's complement arithmetic is always bounded within the unit square, so the global boundedness of the system states is guaranteed. However, this is not the case for the dynamics of SDMs, in which this issue is important for the SDM research community. The reasons for why there is such a difference on the dynamical behaviors of these two systems are that the ways they generate symbolic sequences are different. For SDMs, the symbolic dynamics are generated via the quantization nonlinearity, while the symbolic sequences of digital filters with two's complement arithmetic are generated via the overflow nonlinearity. Although the form of the symbolic dynamical equations for these two systems is exactly the same if the same symbolic sequences are used for expressing the nonlinearity of the systems, they are different if the symbolic sequences are substituted by the corresponding nonlinear functions of the system states.

Risbo reported in 1995 that if one or more than one of the poles of the loop filter are outside the unit circle, then the limit cycle behavior of the SDM will be unstable and chaotic behavior may occur [40]. However, as most of SDMs consist of poles on the unit circle, these results are not commonly applied in many situations.

Closed form analytical expressions were used in [18] to analyze the nonlinear behaviors of SDMs. However, this analysis is based on lots of assumptions and cannot predict complex behaviors.

To analyze the boundedness of the system states of SDMs, the nonlinear quantizer is modelled as a variable gain, in which the gain depends on the ratio of the magnitude of the output to that of the input of the quantizer. This model was first proposed by Ardalan *et. al.* in 1987 [18], and was applied by Stikvoot in 1988 [22] and Baird *et al.* in 1994 [23]. With this model, the boundedness of the system states of the SDM can be derived via an examination of the root locus of the SDM. Although this method is simple, this approach still cannot explain the occurrence of complex behaviors, such as fractal and chaotic behaviors.

In 2002, Feng modelled the SDM as a piecewise discrete-time linear system and performed the stability analysis via deriving a piecewise smooth Lyapunov function [41]. By using a piecewise smooth Lyapunov function, conditions for exhibiting exponentially convergent behavior were derived. However, this result only explained the exponentially convergent behavior, that is, the state variables converge to the origin exponentially for all initial conditions, in which it still cannot explain the occurrence of complex behaviors, such as fractal and chaotic behaviors.

In [19], the analysis of the boundedness of the system states of an SDM is being carried out by Mees and Bergen in 1975 via a large number of tests on the stability of limit cycles, and by studying the stability of limit cycles, conclusion is made on the boundedness of the system states of the SDM. However, the computational complexity is very high because it needs to test a large number of limit cycles. Moreover, the stability of limit cycles does not imply the boundedness of the system states of SDMs.

The invariant set approach is also employed for the stability analysis [24], in which an invariant set is a set that will map to itself under the system mapping. In this thesis, we investigate the feedforward SDM and derive the condition for the existence of an invariant set. It was shown that when an invariant set of the system states exists, a SDM may exhibit chaotic behavior as long as the initial condition is inside this set. However, determining an analytical expression for an invariant set is very difficult. Up to the moment, only a numerical approach is proposed [24] for characterizing an invariant

set. Hence, extensive computer calculation is required. In this research, we determine an analytical expression for an invariant set (Lemma 4, Section 4.2).

Furthermore, Schreier *et. al.* applied an invariant set approach to investigate the boundedness of the system states of an SDM in 1997 [24]. However, the existence of the invariant set only implies the local boundedness of the system states. It does not imply the global boundedness of the system states.

Some methods were proposed to avoid the occurrence of limit cycles. In 1994, Schreier proposed to operate the SDM in a chaotic regime because chaotic signal consists of rich spectrum that breaks the periodic pattern of the output sequence [37]. Another well-known method is to employ a dithering approach to break down the limit cycles [38]. The idea of dithering approach is to inject a noise to the input of the quantizer so that the output bits are flipped and the periodic pattern of the output sequence is broken.

High order SDMs are preferred because they usually perform better noise shaping than lower order SDMs. However, high order SDMs may result to the unboundedness of the system states. The existing control strategies for interpolative SDMs, such as variable structure compensation [42] and time delay feedback control [43], stabilize the loop filter by changing the effective poles of the loop filter. Since the loop filter is usually designed to have a good SNR, the SNR of the controlled SDMs will be affected or even worsen. This may also significantly distort the noise shaping characteristics. Moreover, the parameters in the controller depend on the loop filter parameters, so a particular class of controllers are not able to stabilize all interpolative SDMs. Furthermore, the system states of the controlled SDMs will still be unbounded when the input signal magnitude is further increased, or different initial conditions of the integrator states are employed. In order to control the SDM without changing the effective poles of the loop filter, clipping is employed. Clipping is a simple, common and well-known method that sets the output of the loop filter to a fixed value within an allowable range of the output of the loop filter. This process is able to achieve a bounded loop filter output. However, clipping usually results in limit cycle behavior because the system states are reset to the same value every time. Hence, after certain clock cycles, the same system states will appear for any periodic inputs and result to the occurrence of limit cycles.

In this research, a fuzzy impulsive control strategy is proposed. Fuzzy logic [44] is a logic in which there is no sharp cut between true and false. It allows continuous levels between true and false and it is characterized by fuzzy membership functions. The logic operations among fuzzy variables are based on some fuzzy rules. And these fuzzy rules are formulated based on the heuristic knowledges of the system. Fuzzy control [44] is a control method utilized fuzzy logic and heuristic knowledge. Fuzzy impulsive control was proposed to change the undesirable states of a system to other states by resetting the system states to where the heuristic knowledge determined. The major advantage of employing fuzzy impulsive control is that the state trajectory is guaranteed to be bounded for all initial conditions and limit cycle behaviors can be avoided, no matter what the input signal, initial condition and the filter parameters are. The details of the proposed method, and its advantages, will be discussed in Section 4.1.

According to equation (1.1), the longer the IIR filter, the better the SNR performance can be obtained. However, since the input signal is convolved with the filter, as the order of the filter is increased, the number of multiplications and additions involved in the convolution process is also increased. Hence, this will increase the computation tremendously. For example, for an finite impulse response (FIR) filter with the transition bandwidth equal to 0.1 to achieve ripple magnitude bounded by -40dB, we need 40 coefficients. However, we only need 10 coefficients for the IIR filter case, with 5 coefficients in the numerator and another 5 coefficients in the denominator. Hence, FIR filters are seldom employed in SDMs and we employ IIR filters instead of FIR filters for the optimal design so that we do not need a very long filter length. Design of an optimal IIR filter for an SDM is also challenging and the details will be discussed in Section 5.1.

SDMs are typically designed using Butterworth and Chebyshev filter design rules [45], and optimal designs have been performed based on optimizing operational transconductance amplifier structures [46], speed, resolution and A/D complexity [47] as well as the ratio of peak SNR (PSNR) plus distortion ratio over the power consumption [48] etc. Although these designs have considered many practical issues, the solutions obtained are not globally optimal. It is because the optimization problems involved is not convex and a local optimal solution of a nonconvex problem is not guaranteed to be the global optimal solution, while a local optimal solution of a convex problem is guaranteed

to be the global optimal solution. Here, a convex problem refers to an optimization problem with both the cost function and the corresponding feasible set being convex.

Genetic algorithms have also been applied to perform the optimization [49]. However, the convergence of a genetic algorithm is not guaranteed and the computational complexity of this method is very high. Recently, optimal SDM designs based on comb filter [50] and Laguerre filter [51] structures were proposed. However, the solutions obtained are still sub-optimal ones because structural constraints, such as constraints on the zeros in the impulse response of the comb filter and the poles of the Laguerre filters, are imposed on the designs. In addition, design based on the finite horizon method [52] was proposed, in which this method is to optimize the performance of the system within a finite time support. As this method is only an approximation of the infinite horizon method, the performance is not guaranteed when an infinite time support is considered. Although the approximation can be improved when the length of the horizon window is increased, the computation complexity increases. Other existing optimal design formulation based on practical considerations, such as those reported in [53],[54],[55], are obtained. However, these designs are mainly conducted only based on some simulations without theoretical support.

One way to design rational IIR filters is to firstly initialize a set of the denominator coefficients, and then design the numerator coefficients based on this set of initialized denominator coefficients by solving it as an optimization problem with ripple energy as the cost function and magnitude specifications as the constraints. Then design the denominator coefficients based on the obtained numerator coefficient and iterate these procedures until a converging result is obtained. However, it is not guaranteed that the solution of the iterative procedure will converge [73].

Moreover, the obtained solution depends on the initialization of the denominator coefficients, hence only a local optimal solution can be obtained. Although the divergence problem can be solved by weighting the filter coefficients in each iteration, the frequency characteristics of the filter depend on the weights and the results obtained may be degraded as well. Furthermore, this design method assumes that both the desired magnitude and phase responses of the filter are known. However, as discussed before, sometimes it is difficult to characterize the desired phase response. This can be applied to

Butterworth and Laguerre filter cases because they are nonlinear phase filters. Under this circumstances, the cost function based on the error energy or the absolute error between the desired and designed energy responses will become a fourth order function $\left|\left|H(\omega)\right|^2-\left|H_d(\omega)\right|^2\right|^2$ or a nonsmooth function $\left|\left|H(\omega)\right|^2-\left|H_d(\omega)\right|^2\right|$, where $H(\omega)$ and $H_d(\omega)$ denotes the designed and desired frequency response, respectively. Nevertheless, these problems are not convex.

In this thesis, we design SDMs based on the noise shaping characteristics, the stopband characteristics of the loop filter and the stability conditions for the STF and NTF. The design problem is formulated as two SIP problems. We attempt to apply the dual parameterization approach [56], [57] to solve the problem. It can be shown that global optimal solutions that satisfy the corresponding continuous constraints are guaranteed and a high SNR can be achieved. We further evaluate the performances of the SDMs which are designed based on the SIP approach via various performance indexes, such as the ratio of the total power of the shaped quantization noise to that of the unshaped quantization noise, and the measure of the total loss of channel information capacity after noise shaping, etc. The relationship between the SNR and the OSR, as well as that of the number of bits of the quantizer and the filter order are also investigated.

We can see that most of the existing results on limit cycles are investigated for lowpass SDMs. However, the results for bandpass SDMs are equally important as well. It is because the global boundedness of the system states can be guaranteed and it is easier to characterize the limit cycle behavior so that limit cycle behavior can be avoided in most occasions. In this research, the necessary and sufficient conditions for bandpass SDMs exhibiting limit cycle behavior, as well as the periods and the stability of these limit cycles, are investigated. Moreover, in our investigation, we aim at analyzing and characterizing the occurrence of nonlinear behaviors exhibited in the SDMs so that the SDMs can be designed and controlled to achieve high SNRs and work properly as well.

1.5 Overview of the Thesis

In our investigation, we find that elliptic fractal patterns do not only occur in single bit bandpass SDMs, but also occur in multi-bit bandpass SDMs, for the case when

the saturation regions of the multi-bit quantizers are not activated and a large number of bits are used for the implementation of the quantizers. Moreover, we find that the visual appearance of the phase portraits of the infinite state machine and the finite state machine with high bit quantizers can be very different. These phenomena are different from those previously reported for the digital filter with two's complement arithmetic and some more interesting phenomena are explored.

It has been found that a class of bandpass SDMs will exhibit fractal patterns in the phase plane when the system matrices are marginally stable. For the case when the system matrices are strictly stable, we also found that near chaotic phenomenon would occur. This phenomenon is not very intuitive because for a bandpass SDM with stable closed loop filter, the system state is expected to diverge. In this research, we found from the phase portraits that elliptic fractal pattern confined in two trapezoidal region would occur.

Existing control strategies, such as variable structure compensation and time delay feedback control, have some drawbacks. A critical one is that these control strategies change the effective poles of the loop filters, which implies that the system states of the SDM will still be unbounded when the magnitude of the input signal is increased. Some nonlinear control strategies, such as the clipping method, were proposed. However, these control strategies usually result to the occurrence of limit cycle behavior because the system states are always reset to the same values. In this research, we develop a control strategy based on fuzzy impulsive control technique. This control strategy is to change the system states directly instead of changing the effective poles of the loop filters. Hence, the system states can still be controlled and bounded even though the magnitude of the input signal is increased. Note that the system states are reset to different values, so limit cycle behavior does not occur.

Existing sub-optimal designs based on comb filter, Butterworth filter, Chebyshev filter and Laguerre filter were proposed. However, they assume certain structures on the filters. For example, all poles of Laguerre filter are the same, while those of Butterworth filter are on the same circle in the complex plane, and there are many zeros in the impulse response of the comb filter. Performance of an SDM may be improved if these filter structures are relaxed. In this research, we formulate the design problem as two

optimization problems. The first optimization problem is to minimize the passband energy of the denominator of the loop filter transfer function (excluding the DC poles), subject to the continuous constraint on the maximum modulus square of the denominator of the loop filter transfer function. This cost function is chosen because small passband energy of the denominator of the loop filter transfer function corresponds to low NTF and almost unity gain STF which results to good SNR. The second optimization problem is to minimize the stopband energy of the numerator of the loop filter transfer function, subject to the continuous constraint on the stability condition of the NTF and STF. This cost function is chosen because low stopband energy of the numerator of the loop filter transfer function corresponds to good frequency selectivity of the loop filter. The optimization problems are actually quadratic semi-infinite programming (SIP) problems. By employing the dual parameterization approach, global optimal solutions that satisfy the corresponding continuous constraints can be guaranteed if the filter length is sufficiently long. The advantages of this formulation are the guarantee of the stability of the NTF and STF, applicability to the design of rational IIR filters without imposing specific filter structures, and avoidance of iterative designs of the numerator and denominator coefficients. Our simulation results show that this design can yield a significant improvement in SNR and has a larger input range for bounded system states, compared to the existing designs.

CHAPTER II. ELLIPTIC FRACTAL PATTERNS IN MULTI-BIT SDMS

One reason why the research and development of SDMs has been concerned with the use of multi-bit quantizers is that the system states of the SDMs with single bit quantizers are usually unbounded, typically when the inputs are overloaded [65]. This is because when the input is seriously overloaded, the signal distortion and the nonlinear effects will be significant. Consequently, the SNR would be degraded. In this chapter, bandpass SDMs are investigated. The reasons why a bandpass SDM is investigated were stated in Section 1.3.3. As discussed in Section 1.3.1 that even for the class of bandpass SDMs with a single bit quantizer, they could exhibit state space dynamics represented by elliptic fractal patterns confined within two trapezoidal regions [25]. The question arises whether similar patterns will occur for the multi-bit cases. If the saturation region of a quantizer is not activated and there is infinite number of bits for the implementation of the quantizer, then the bandpass SDM will become linear system and fractal behavior will not occur. Consequently, one may ask when the number of bits of the quantizer is increased, but the saturation region is still not activated, will the nonlinear behavior never occur? If not, what behavior will be shown on the phase portrait as the number of bits of the quantizer is increased?

2.1 System description

Consider the structure of the SDM shown in Figure 1.4, suppose that the loop filter is a second order bandpass filter with the following transfer function:

$$F(z)=\frac{2\cos\theta z^{-1}-z^{-2}}{1-2\cos\theta z^{-1}+z^{-2}}.$$

As opposed to standard lowpass SDM systems, bandpass SDMs are usually designed to operate on high frequency narrowband signals by shaping the noise away from that frequency, denoted as f_0 [33] and $\theta=\frac{2\pi f_0}{f_s}$, in which f_s denotes the sampling frequency. Assume that $\theta\in(-\pi,\pi)\setminus\{0\}$. When $\theta\in\{-\pi,0,\pi\}$, the system is either a lowpass or highpass SDM, which is out of the scope of our investigation. We also assume that the

input signal $u(k)$ is real signal. Then at the desired frequency $z = e^{\pm j\theta}$, it can be easily checked that

$$\begin{aligned}
\text{NTF} &= \left.\frac{1}{1+F(z)}\right|_{z=e^{\pm j\theta}} \\
&= \left.\frac{1}{1+\dfrac{2\cos\theta z^{-1}-z^{-2}}{1-2\cos\theta z^{-1}+z^{-2}}}\right|_{z=e^{\pm j\theta}} \\
&= 1-2\cos\theta z^{-1}+z^{-2}\Big|_{z=e^{\pm j\theta}} \\
&= 0
\end{aligned}$$

because $e^{\pm j\theta}$ are the poles of the filter, and

$$\begin{aligned}
\text{STF} &= \left.\frac{F(z)}{1+F(z)}\right|_{z=e^{\pm j\theta}} \\
&= \left.\frac{\dfrac{2\cos\theta z^{-1}-z^{-2}}{1-2\cos\theta z^{-1}+z^{-2}}}{1+\dfrac{2\cos\theta z^{-1}-z^{-2}}{1-2\cos\theta z^{-1}+z^{-2}}}\right|_{z=e^{\pm j\theta}} \\
&= 2\cos\theta z^{-1}-z^{-2}\Big|_{z=e^{\pm j\theta}} \\
&= 1-1+2\cos\theta z^{-1}-z^{-2}\Big|_{z=e^{\pm j\theta}} \\
&= 1-\left(1-2\cos\theta z^{-1}+z^{-2}\right)\Big|_{z=e^{\pm j\theta}} \\
&= 1
\end{aligned}$$

Note that this system is also studied by Feely in [33].

Since the input-output relationship of the filter is governed by its transfer function, we have

$$F(z) = \frac{Y(z)}{U(z)-S(z)},$$

where $Y(z)$ and $S(z)$ are the z-transform of the output of the loop filter and the quantizer, respectively. Expressing this as a difference equation, we have:

$$y(k)-2\cos\theta y(k-1)+y(k-2) = 2\cos\theta\big(u(k-1)-s(k-1)\big)-\big(u(k-2)-s(k-2)\big),$$

which further implies that

$$y(k) = 2\cos\theta\big(u(k-1)-s(k-1)\big)-\big(u(k-2)-s(k-2)\big)+2\cos\theta y(k-1)-y(k-2).$$

By writing this equation as matrix form, we have:

$$\begin{bmatrix} y(k-1) \\ y(k) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 2\cos\theta \end{bmatrix} \begin{bmatrix} y(k-2) \\ y(k-1) \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ -1 & 2\cos\theta \end{bmatrix} \left(\begin{bmatrix} u(k-2) \\ u(k-1) \end{bmatrix} - \begin{bmatrix} s(k-2) \\ s(k-1) \end{bmatrix} \right).$$

Define the state variables of the SDM as the delayed versions of the output of the loop filter, that is $\mathbf{x}(k) \equiv [x_1(k) \quad x_2(k)]^T \equiv [y(k-2) \quad y(k-1)]^T$, in which the superscript ^{T} denotes the transpose operator. Denote $\mathbf{u}(k)$ as a vector containing the past two consecutive points from the input signal $u(k)$, that is $\mathbf{u}(k) \equiv [u(k-2) \quad u(k-1)]^T$, and denote $\mathbf{s}(k)$ as a quantized system state, that is $\mathbf{s}(k) \equiv Q(\mathbf{x}(k)) \equiv [Q(x_1(k)) \quad Q(x_2(k))]^T$, where Q is a mid-rise quantizer and represented as

$$Q(y) \equiv \begin{cases} \dfrac{y}{|y|} & |y| > L \\ 0 & y = 0 \\ \dfrac{y\Delta}{|y|} \text{ceil}\left(\dfrac{|y|}{\Delta} \right) & |y| \le L \text{ and } y \ne 0 \end{cases},$$

in which $|y|$ denotes the absolute value of y, $\text{ceil}(y)$ denotes the nearest integer of y towards infinity, Δ denotes the step size of the quantizer and L denotes the saturation level of quantizer. The relationship between Δ and L is governed by

$$\Delta \equiv \frac{L}{2^{N-1}},$$

where N denotes the number of bits of the quantizer. Figure 2.1 shows an example of the input-output characteristics of this quantizer with $N = 3$.

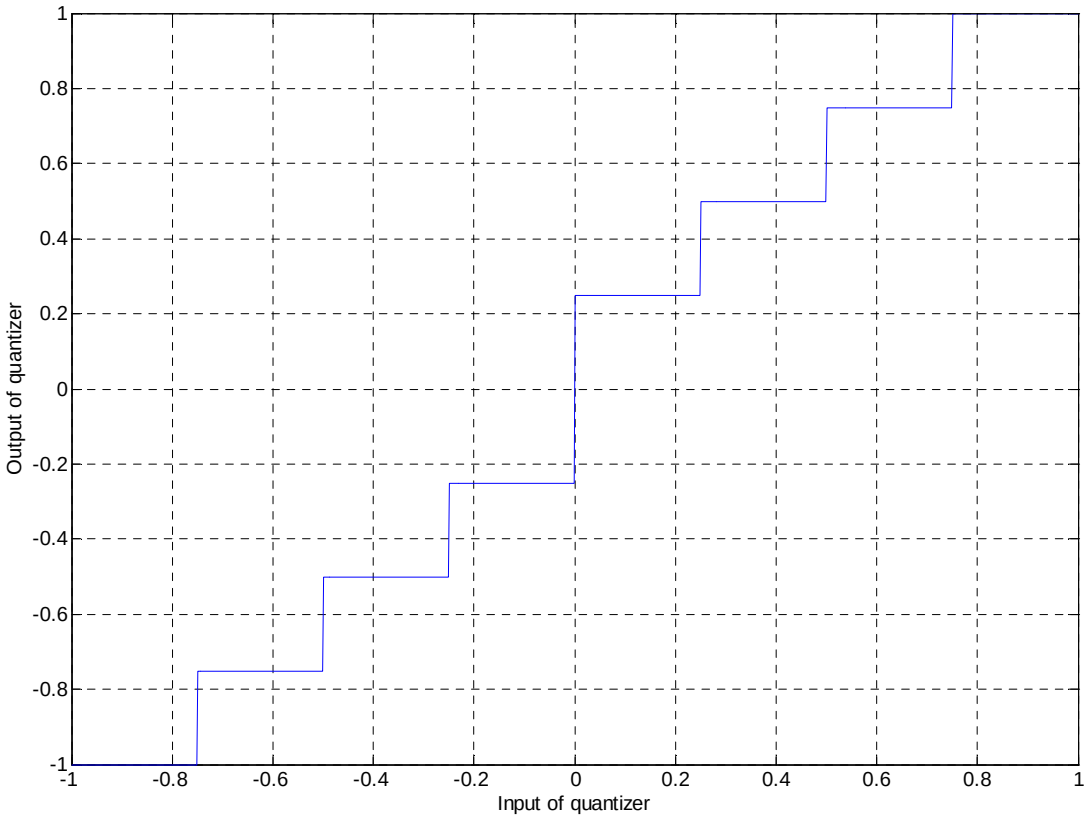


Figure 2.1 Input-output characteristics of $Q_{mid-rise}(y)$ with $N=3$.

Note that the quantizer we employed has the quantization jump at the origin being double than that at the other discontinuous points. Compared to the conventional mid-rise quantizers with $N>1$

$$Q_{mid-rise}(y)=\begin{cases}\dfrac{|y|\Delta}{y}\left(\text{ceil}\left(\dfrac{|y|}{\Delta}\right)-\dfrac{1}{2}\right) & |y|\le\left(2^{N-1}-1\right)\Delta \\ \dfrac{|y|\Delta}{y}\left(2^{N-1}-\dfrac{1}{2}\right) & |y|>\left(2^{N-1}-1\right)\Delta\end{cases},$$

and the conventional mid-thread quantizers with $N>1$

$$Q_{mid-thread}(y)=\begin{cases}\dfrac{|y|\Delta}{y}\text{round}\left(\dfrac{|y|}{\Delta}\right) & |y|<\left(2^{N-1}-\dfrac{3}{2}\right)\Delta \\ \dfrac{|y|\Delta}{y}\left(2^{N-1}-1\right) & |y|\ge\left(2^{N-1}-\dfrac{3}{2}\right)\Delta\end{cases},$$

where $\text{round}(y)$ denotes the rounding operator, these two commonly employed quantizers have uniform quantization at the origin. We employ this quantizer because this

quantizer is employed in [66]. As our main objective is on the investigation of the occurrence of the fractal behaviors of multi-bit SDMs, in order to have a fair comparison, a quantizer with the same nonlinearity as that in [66] should be employed.

The bandpass SDM can be described by the following state space equation:

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{B}(\mathbf{u}(k) - \mathbf{s}(k)) \text{ for } k \geq 0, \tag{2.1}$$

where

$$\mathbf{A} \equiv \begin{bmatrix} 0 & 1 \\ -1 & 2\cos\theta \end{bmatrix} \text{ and } \mathbf{B} \equiv \begin{bmatrix} 0 & 0 \\ -1 & 2\cos\theta \end{bmatrix}. \tag{2.2}$$

Since $\mathbf{s}(k)$ is a vector containing discrete output sequences, the values of $\mathbf{s}(k)$ can be viewed as symbols and $\mathbf{s}(k)$ is called a symbolic sequence.

We assume that $u(k)$ is a constant input, that is $\mathbf{u}(k) = \mathbf{u}$ for $k \geq 0$, as explained in detail in Section 1.3.3. Once the initial condition $\mathbf{x}(0)$, the filter parameter of the system θ, the input step size $\mathbf{u}$ and the number of bits of the quantizer N are given, the system state vector $\mathbf{x}(k)$ and the symbolic sequence $\mathbf{s}(k)$ (or the output sequence of the system) can be uniquely determined by equation (2.1).

2.2 Nonlinear behaviors of multi-bit SDMs

Figure 2.2a-2.2c shows the phase portraits of a bandpass SDM [25] and Figure 2.2d-2.2f shows the phase portraits of the bandpass SDM shifted both horizontally and vertically by 0.4 with

$$\theta = \cos^{-1}(-0.158532),$$

$$\mathbf{u} = -0.3[1 \quad 1]^T,$$

$$\mathbf{x}(0) = [0 \quad 0.5]^T,$$

$$L = 1$$

and different values of N. We plot the shifted phase portraits because the difference between the linear and nonlinear systems can be demonstrated more clearly. The values of the state variables are bounded by -1 and 1. This implies that the system is operating in the quantization region, that is:

$$|x_i(k)| \leq L \text{ for } k \geq 0 \text{ and for } i = 1,2,$$

and the saturation regions ($|x_i(k)| > L$) are not activated. Therefore, as the number of bits of the quantizer is increased, the bandpass SDM becomes a closer approximation of the corresponding linear system. To analyze the corresponding behavior of the linear system, the steady state value of the corresponding linear system is

$$\begin{aligned}
&\lim_{z\to 1}(1-z^{-1})S(z)\Big|_{NTF=0} = \lim_{z\to 1}(1-z^{-1})U(z)STF(z) \\
&= \lim_{z\to 1}(1-z^{-1})U(z)\frac{F(z)}{1+F(z)} \\
&\approx 0.4
\end{aligned} \quad .$$

So it is expected that the system converges to a fixed point located at $[0.4 \quad 0.4]^T$. However, elliptic fractal patterns are exhibited on the phase portrait even when a large number of bits (N=37) are used for the implementation of the quantizers, as shown in Figure 3.2f.

Recalling the simulation results in [27], the visual appearance of the phase portraits between the infinite state machine and the finite state machine with high bit quantizers are different. This result is different from the existing results on second order digital filters with two's complement arithmetic [66], where there are visually indistinguishable elliptic fractal patterns shown on the phase portraits when 16 bits are used for the implementation of the quantizer. Besides, the fractal pattern may occur for low bit quantizers as shown in [25]. This result is also different from the existing results [66], in which the fractal behavior exists only for high bit quantizers.

It was found that most initial conditions would lead to elliptic fractal behaviors while the others would lead to limit cycle or chaotic behaviors. Figure 2.3a-2.3c shows the phase portraits of a bandpass SDM [25] with

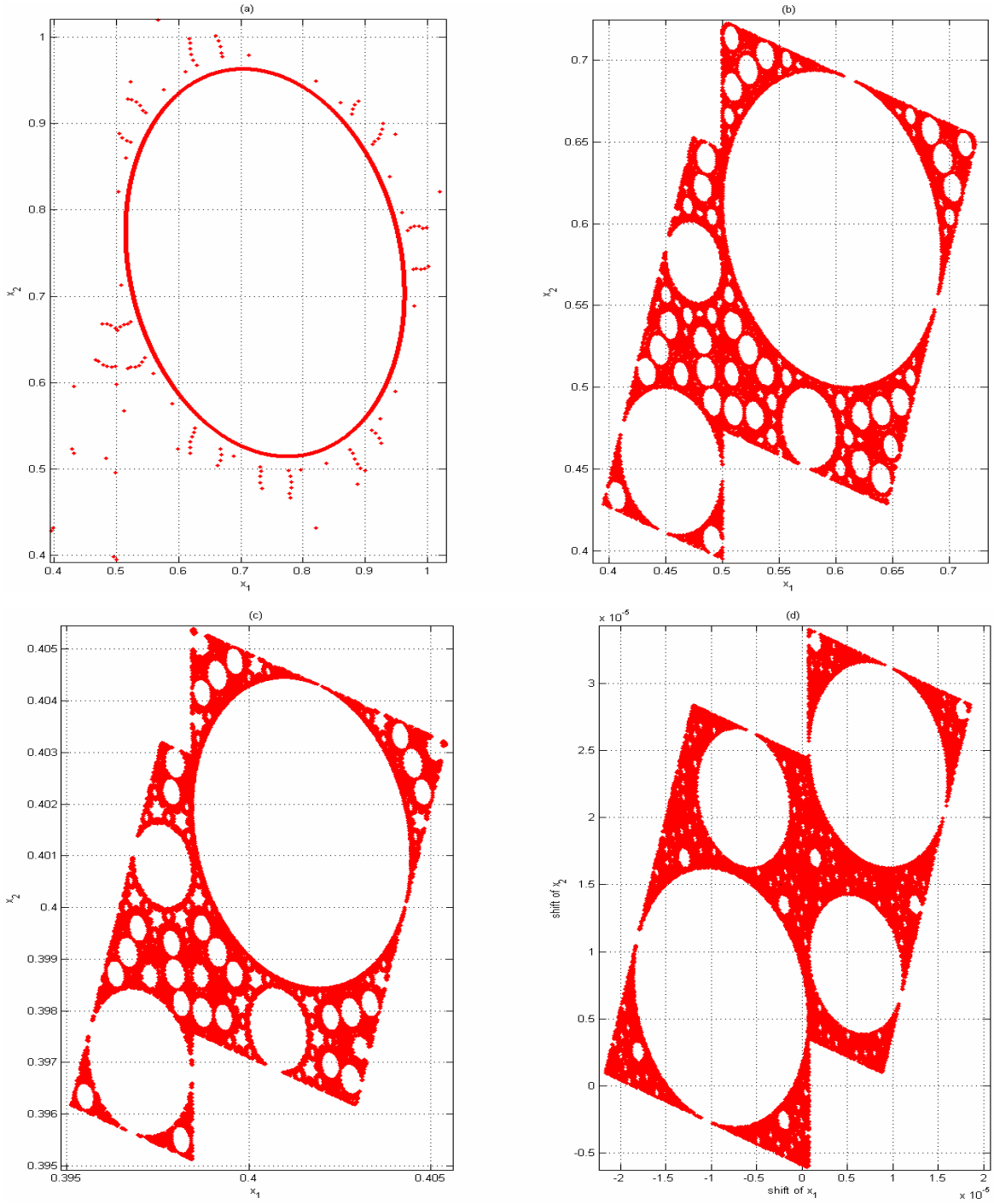
$$\theta = \cos^{-1}(-0.158532),$$

$$\mathbf{u} = -0.3[1 \quad 1]^T,$$

$$\mathbf{x}(0) = [1 \quad 0]^T,$$

$$L = 1$$

and different values of N.


(a)
(b)
(c)
(d)
x1
x2
shift of x1
shift of x2
x 10^-5

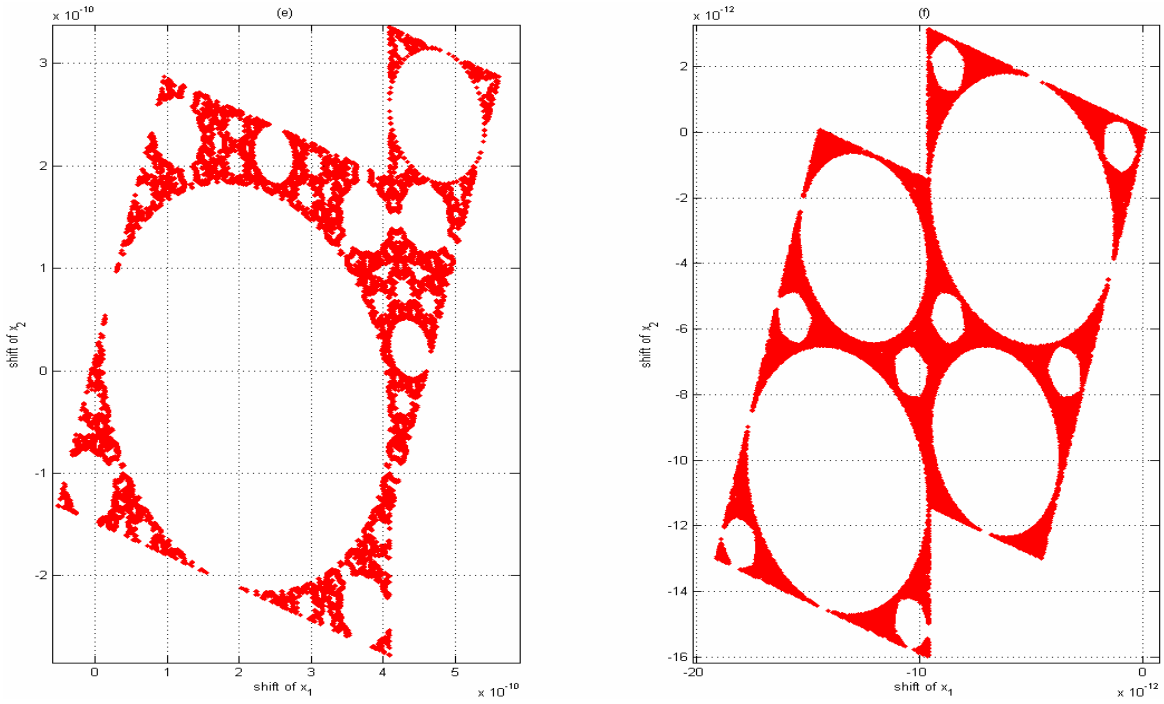


Figure 2.2 The phase portraits of bandpass SDMs (a) $N=2$ (b) $N=3$. (c) $N=8$. The shifted phase portraits of bandpass SDMs. (d) $N=16$. (e) $N=32$. (f) $N=37$.

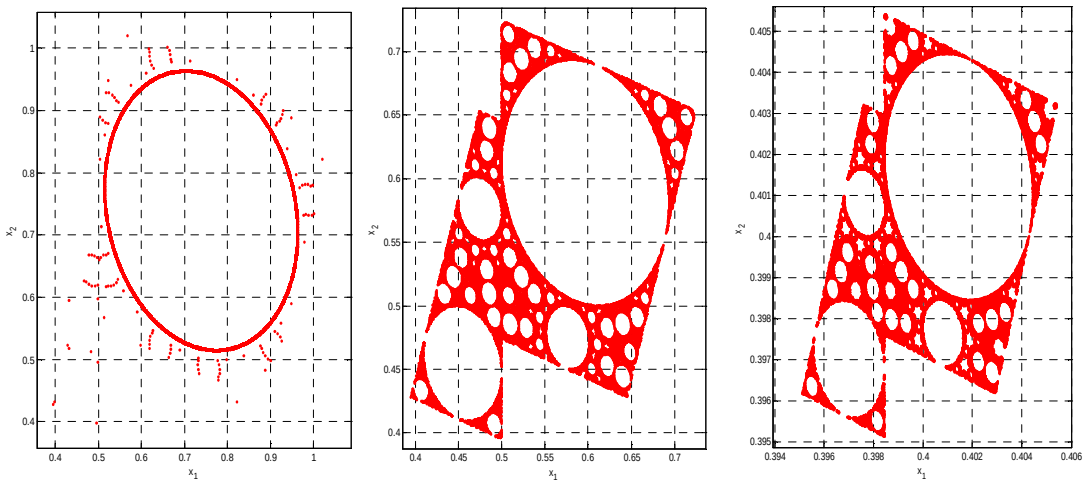


Figure 2.3 The phase portraits of bandpass SDMs with different initial conditions (a) $N=2$ (b) $N=3$. (c) $N=8$.

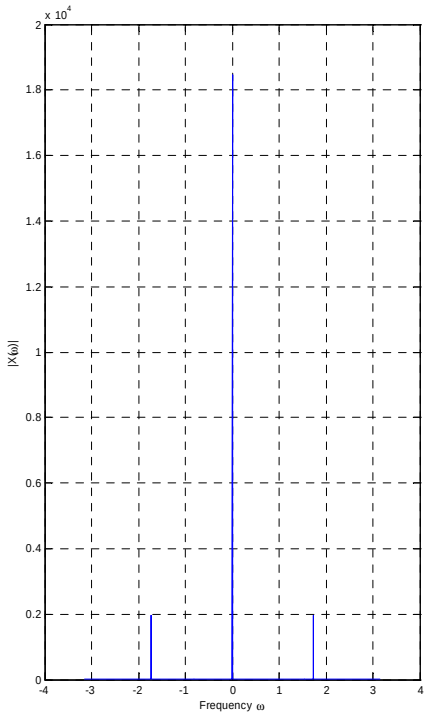
Moreover, some interesting results are found. It is shown in Figure 2.2 that the trajectories converge to a single ellipse when $N=2$, while the trajectories exhibit elliptic fractal patterns when $N=3$, $N=8$, $N=16$, $N=32$ and $N=37$. When the number of bits is increased by one, such as from $N=2$ to $N=3$, the trajectory will change

dramatically from a single ellipse to a trapezoidal elliptic fractal pattern. This phenomenon demonstrates that the system is very sensitive to the nonlinearity.

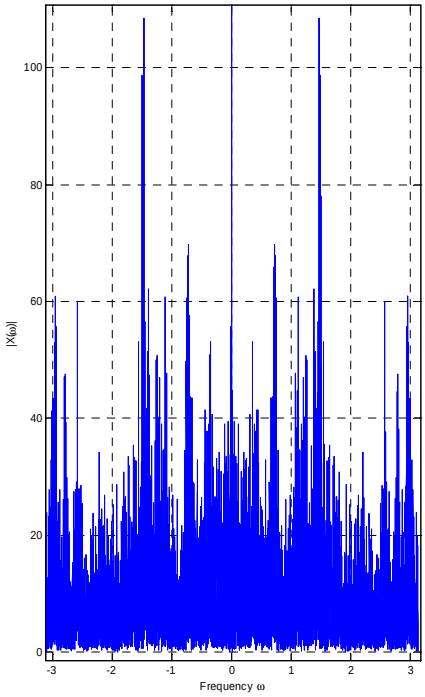
Besides, according to [27] that the elliptic fractal pattern is expected to be influenced by the value of θ, our results show that the complexity of the nonlinear behaviors of the SDM depend on the number of bits of the quantizer as well. When the number of bits of the quantizer increases progressively, the overall size of the elliptic fractal pattern will diminish in which the exact relationship is too complicated to be analyzed, because the relative size of the trapezoids, the fractal pattern and the resolution of the fractal pattern will vary irregularly.

Figure 2.4 shows the spectra of the state variables. It can be shown in Figure 2.4 (a) that there are two impulses in the frequency spectrum. That means, there are only two frequency components. This implies that the symbolic sequence is periodic which corresponds to limit cycle behavior.

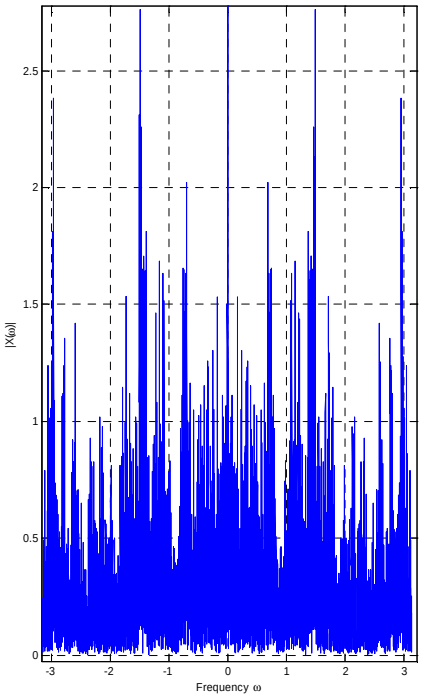
Figure 2.4 (b) and Figure 2.4 (c) shows that the frequency spectra are very rich. From the figures, we can see that the spectra are continuous and the symbolic sequences are aperiodic. This means that the SDMs do not exhibit limit cycle behavior. This result is important because it implies that limit cycle can be avoided for the multi-bit cases simply by operating them in fractal behavior regime.


x 10^4
|X(ω)|
Frequency ω

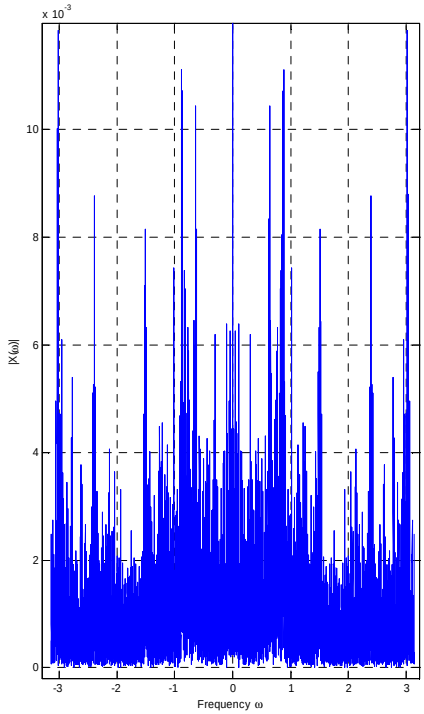
(a)


|X(ω)|
Frequency ω

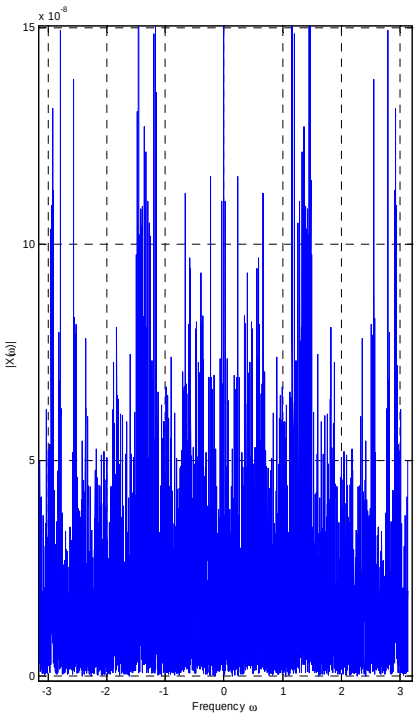
(b)


|X(ω)|
Frequency ω

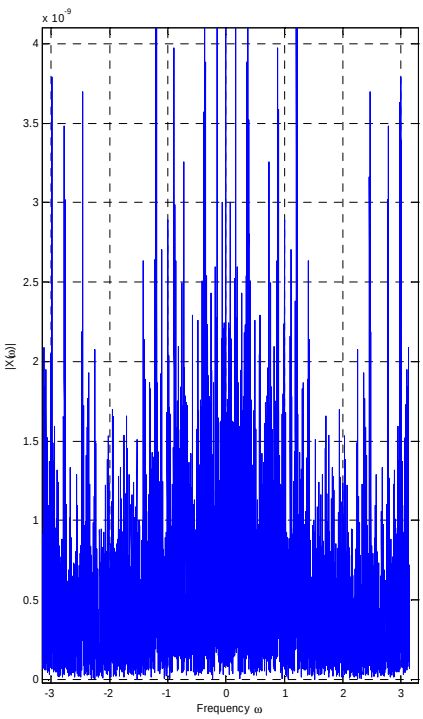
(c)


x 10^-3
|X(ω)|
Frequency ω

(d)


x 10^-8
|X(ω)|
Frequency ω

(e)


x 10^-9
|X(ω)|
Frequency ω

(f)

Figure 2.4 The frequency spectra of the bandpass SDMs. (a) $N = 2$ (b) $N = 3$. (c) $N = 8$. (d) $N = 16$. (e) $N = 32$. (f) $N = 37$.

2.3 Conclusions

In this chapter, we have showed that elliptic fractal patterns would also occur in bandpass SDMs with multi-bit quantizers, for the case when the saturation regions of the multi-bit quantizers are not activated, and a large number of bits are used for the implementation of the quantizers. Moreover, the visual appearance of the phase portraits between the infinite state machines and the finite state machines with high bit quantizers are different. These phenomena are different from those reported for the digital filter with two's complement arithmetic [66], as summarized in Section 1.3.4. Moreover, a bit change of the quantizer can result in a dramatic change in the phase portrait. When the trajectories of the corresponding linear system converge to a fixed point, the regions of the elliptic fractal pattern will diminish in size and approach to the fixed point as the number of bits of the quantizers increases. However, since the multi-bit SDM exhibits fractal behavior, this implies that the limit cycle behavior can be avoided by operating the multi-bit SDM in the fractal or chaotic behavior regime.

CHAPTER III. NONLINEAR BEHAVIORS OF SDMS WITH STABLE SYSTEM MATRICES

As discussed in Section 1.4, some researchers utilize the nonlinear behavior generated by the quantizer in SDMs to suppress unwanted tones [35]. The simplest existing method is to via the root locus approach. The corresponding linearized system is operated in the unstable region. The objective is to operate the system in a chaotic regime so that the rich spectra of these chaotic output signals break down the dominant oscillations at the outputs. However, operating the corresponding linearized system in an unstable region would cause the system states of the SDM to be unbounded.

In practical situation, that is, for those SDMs consisting of integrators and being employed in A/D conversion [36], there are leakages on these integrators. This originates from the internal resistances of the components. Even though the leakages may sometimes be negligible, engineers and circuit designers may impose leakages on the integrators as another method to guarantee the boundedness of the system states of the SDMs.

It is well known that even though the linearized closed loop SDM is stable, it does not guarantee that the system states of the SDM are bounded, and vice versa. Therefore, the boundedness of the system states cannot be achieved via the root locus method. Consider the following example. Suppose that the loop filter transfer function is $F(z)=\dfrac{2rz^{-1}-r^2z^{-2}}{1-2rz^{-1}+r^2z^{-2}}$, and the transfer function of the feedback control system is $C(z)=K$, where K and r are constants, then the transfer function of the linearized closed loop SDM is

$$T(z)=\frac{F(z)}{1+KF(z)}=\frac{\dfrac{2rz^{-1}-r^2z^{-2}}{1-2rz^{-1}+r^2z^{-2}}}{1+K\dfrac{2rz^{-1}-r^2z^{-2}}{1-2rz^{-1}+r^2z^{-2}}}=\frac{2rz^{-1}-r^2z^{-2}}{1-2rz^{-1}+r^2z^{-2}+K\left(2rz^{-1}-r^2z^{-2}\right)}.$$

If $0<r<1$, then $F(z)$ will be strictly stable. When $K<0$ and $r\to 1^-$, then $T(z)$ will be unstable because the poles will be outside the unit circle. However, as we know that the output of the SDM is always bounded because of the quantizer, any constant (even

though it is negative) multiplied to the output of the SDM plus the input of the SDM is bounded if the input of the SDM is bounded. That means, the input of the loop filter is bounded. Since $F(z)$ is strictly stable, the output of the loop filter has to be bounded even though $T(z)$ is unstable. Hence, the instability of the linearized closed loop SDM does not imply the unboundedness of the state variables of the SDM.

Similarly, if $r=1$, then $F(z)$ will be marginally stable and $T(z)$ will be unstable for $K<0$. Suppose that the input of the SDM and the output of symbolic sequences do not contain a frequency component exactly equal to the natural frequency of the loop filter, then the output of the loop filter will be bounded if the input of the SDM is bounded. This also shows that the instability of the linearized closed loop SDM does not imply the unboundedness of the state variables of the SDM.

Lastly, if $r>1$, then $F(z)$ is unstable. However, by using the root locus technique, we show that $\exists K>0$ such that $T(z)$ is stable. That means, the output of the loop filter may be unbounded for some bounded input of the SDM because the loop filter is unstable. This shows that the stability of the linearized closed loop SDM does not imply the boundedness of the state variables of the SDM. Figure 3.1a, Figure 3.1b and Figure 3.1c show the pole zero plots of the loop filters which are strictly stable, marginally stable and unstable, respectively. Figure 3.1d, Figure 3.1e and Figure 3.1f show the pole zero plots of the linearized closed loop SDMs with loop filters being strictly stable, marginally stable and unstable, respectively. Figure 3.1g, Figure 3.1h and Figure 3.1i show the output of the loop filters which are strictly stable, marginally stable and unstable, respectively, when the input of the SDM being a step signal of step size 0.9 and zero initial condition. It can be seen from Figure 3.1g and Figure 3.1h that the loop filter responses are bounded even though $T(z)$ is unstable, and it can be seen from Figure 3.1i that the loop filter response is unbounded even though $T(z)$ is stable.

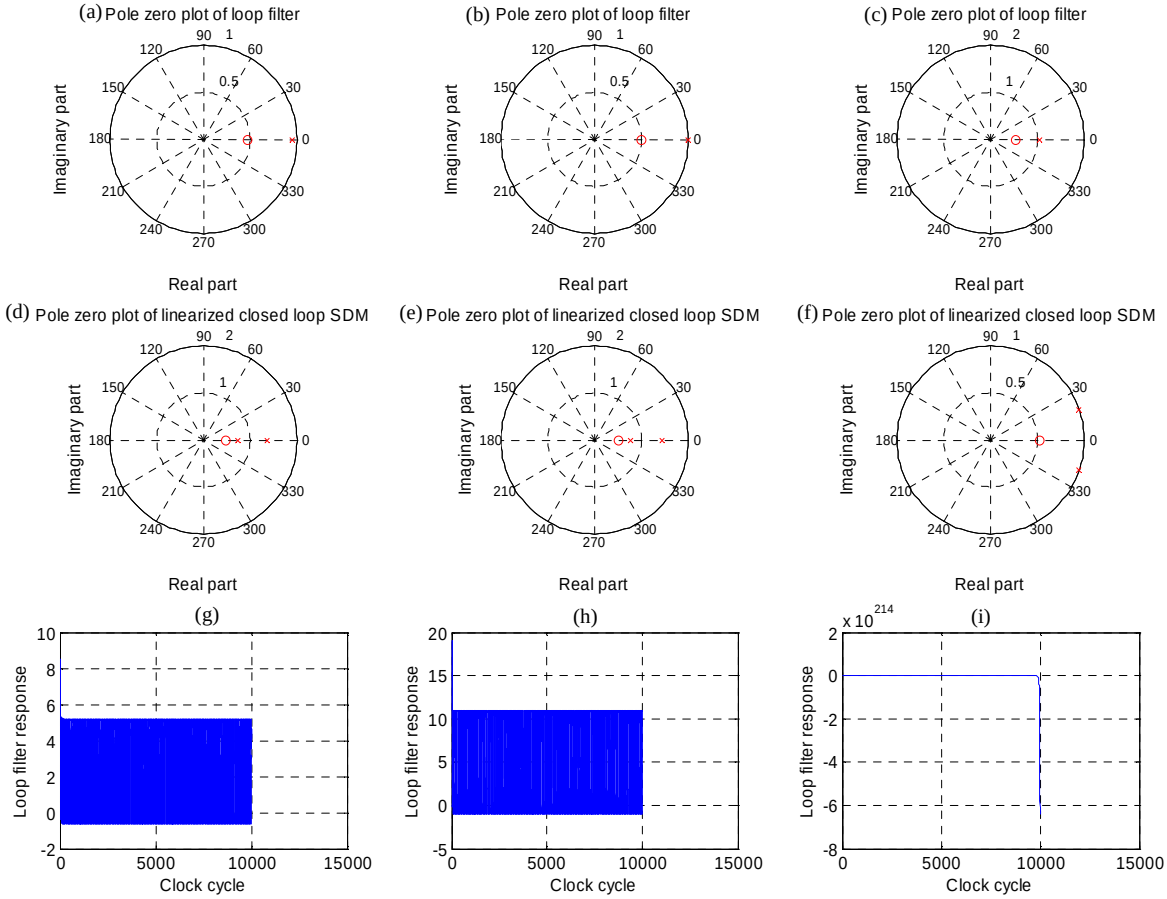


Figure 3.1 Pole zero plots for both the open loop and closed loop linearized transfer functions and the responses of the loop filters for the strictly stable, marginally stable and unstable loop filters cases.

In this chapter, we only consider the case when the loop filter is strictly stable. It is worth noting that there are some analytical results on the bandpass SDMs, for example, [25], [33] and [34], but most analyses are based on marginally stable system matrices only. For the bandpass SDMs with strictly stable system matrices, the existing results are primarily concerned with limit cycles with short periods, but not with near fractal or near chaotic behavior [77]. Theoretically, systems with strictly stable system matrices will cause the trajectories to converge to some fixed points, so real fractal and real chaotic behaviors would not occur.

In our investigation, it is found that near fractal or near chaotic patterns [77] will be exhibited in the phase plane when the system matrices are strictly stable. This occurs when the period of the limit cycles are very large that the difference of the phase portraits

between the near fractal and the real fractal behaviors [77], or that between the near chaotic and the real chaotic behaviors, are visually indistinguishable. Based on the derived analytical results, some interesting results are found. If the bandpass SDM exhibits periodic output, then the period of the symbolic sequence will be equal to the limiting period of the state variables, where the limiting period is defined as the period when the steady state of the symbolic sequence is reached. Second, if the system states converge to some fixed points on the phase portrait, these fixed points will depend directly on the corresponding symbolic sequences, in which they will depend indirectly on the initial condition.

3.1 State space formulation

Consider the structure of the SDM shown in Figure 1.3, suppose that the loop filter is a second order strictly stable bandpass filter with the following transfer function:

$$F(z)=\frac{2r\cos\theta z^{-1}-r^2z^{-2}}{1-2r\cos\theta z^{-1}+r^2z^{-2}},$$

where $0<r<1$. It is worth noting that sum of the numerator and the denominator polynomials of this loop filter transfer function is equal to one, this implies that the linearized closed loop SDM is stable. This system is also studied by Feely in [33]. Since the poles of the filter are $re^{j\theta}$ and $re^{-j\theta}$, the corresponding magnitude is $\left|re^{j\theta}\right|=\left|re^{-j\theta}\right|=r<1$, which is strictly inside the unit circle, this case refers to the strictly stable case. The leakage of the system depends on the values of r. If r is closer to 0, then the poles will be closer to the origin and the leakage is more serious. If r is closer to 1, then the poles will be closer to the unit circle and the leakage will be less significant. For an ideal lossless bandpass SDM, $r=1$, the system reduces to that described in [33], which is marginally stable. Here, we also assume that $\theta\in(-\pi,\pi)\setminus\{0\}$ and the input signal $u(k)$ is a real signal. These assumptions are also made in Section 2.1. Using a similar approach discussed in Section 2.1, we have:

$$y(k)-2r\cos\theta y(k-1)+r^2y(k-2)=2r\cos\theta(u(k-1)-s(k-1))-r^2(u(k-2)-s(k-2)),$$

which further implies that

$$y(k)=2r\cos\theta(u(k-1)-s(k-1))-r^2(u(k-2)-s(k-2))+2r\cos\theta y(k-1)-r^2y(k-2),$$

where $y(k)$ and $s(k)$ are the outputs of the loop filter and the quantizer, respectively. By writing this equation into matrix form, we have:

$$\begin{bmatrix} y(k-1) \\ y(k) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -r^2 & 2r\cos\theta \end{bmatrix} \begin{bmatrix} y(k-2) \\ y(k-1) \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ -r^2 & 2r\cos\theta \end{bmatrix} \left(\begin{bmatrix} u(k-2) \\ u(k-1) \end{bmatrix} - \begin{bmatrix} s(k-2) \\ s(k-1) \end{bmatrix} \right).$$

Similarly, denote

$\mathbf{x}(k) \equiv [x_1(k) \quad x_2(k)]^T \equiv [y(k-2) \quad y(k-1)]^T$ as the system state of the system, $\mathbf{u}(k) \equiv [u(k-2) \quad u(k-1)]^T$ as a vector containing the past two consecutive points from the input signal $u(k)$,

$$\mathbf{A} \equiv \begin{bmatrix} 0 & 1 \\ -r^2 & 2r\cos\theta \end{bmatrix}$$

as the system matrix of the system,

$$\mathbf{B} \equiv \begin{bmatrix} 0 & 0 \\ -r^2 & 2r\cos\theta \end{bmatrix}$$

as the matrix associated with the nonlinearity and the input, and

$$\mathbf{s}(k) \equiv [Q(x_1(k)) \quad Q(x_2(k))]^T \text{ for } k \geq 0,$$

in which

$$Q(y) \equiv \begin{cases} 1 & y \geq 0 \\ -1 & \text{otherwise} \end{cases}.$$

Note that the quantizer has two levels, in which this single bit quantizer is the most common quantizer employed for SDMs. Then we have

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{B}(\mathbf{u}(k) - \mathbf{s}(k)) \text{ for } k \geq 0. \tag{3.1}$$

Since

$$\mathbf{s}(k) \in \{[1 \quad 1]^T, [1 \quad -1]^T, [-1 \quad 1]^T, [-1 \quad -1]^T\} \text{ for } k \geq 0, \tag{3.2}$$

the value of $\mathbf{s}(k)$ can be viewed as symbols, and $\mathbf{s}(k)$ is called a symbolic sequence.

In this case, we also assume that the input is DC. That is, $\mathbf{u}(k) = \mathbf{u}$ for $k \geq 0$, as explained in Section 1.3.3.

3.2 Limit cycle behaviors

In this Section, the limit cycle behavior of the SDM will be analyzed.

Note that $\mathbf{A}$ is a full rank matrix because $r \neq 0$. The eigen decomposition of $\mathbf{A}$ is

$$\mathbf{A} = \mathbf{TDT}^{-1}, \tag{3.3}$$

where the diagonal elements of $\mathbf{D}$ and the columns of $\mathbf{T}$ are the eigenvalues and the eigenvectors of $\mathbf{A}$, respectively. It is worth noting that $\mathbf{T}^{-1}$ exists because $\theta \notin \{-\pi, 0, \pi\}$. Since the poles of the filters are $re^{j\theta}$ and $re^{-j\theta}$, we have:

$$\mathbf{D} \equiv \begin{bmatrix} re^{j\theta} & 0 \\ 0 & re^{-j\theta} \end{bmatrix} \tag{3.4}$$

and

$$\mathbf{T} \equiv \begin{bmatrix} \frac{1}{\sqrt{r}} e^{-\left(\frac{j\theta}{2}\right)} & \frac{1}{\sqrt{r}} e^{\frac{j\theta}{2}} \\ \sqrt{r} e^{\frac{j\theta}{2}} & \sqrt{r} e^{-\left(\frac{j\theta}{2}\right)} \end{bmatrix}. \tag{3.5}$$

Let M be the period of the steady state of the output sequences (if it exists), that is

$$\mathbf{s}(k_0 + M + i) = \mathbf{s}(k_0 + i) \ \ \forall i \geq 0, \tag{3.6}$$

in which $M \in Z^+$ and $k_0 \in Z^+ \bigcup \{0\}$ such that $s(k)$ reaches the steady states. Define

$$\mathbf{x}_0^* \equiv \sum_{n=0}^{M-1} \mathbf{T}\mathbf{D}^{M-1-n} \left(\lim_{p \to +\infty} \sum_{m=0}^{p-1} \mathbf{D}^{mM} \right) \mathbf{T}^{-1} \mathbf{B}(\mathbf{u} - \mathbf{s}(k_0 + n)) \tag{3.7}$$

and

$$\mathbf{x}_i^* \equiv \mathbf{A}^i \mathbf{x}_0^* + \sum_{m=0}^{i-1} \mathbf{A}^{i-1-m} \mathbf{B}(\mathbf{u} - \mathbf{s}(k_0 + m)) \text{ for } i = 1,2,\cdots,M-1. \tag{3.8}$$

In the following lemma we show that $\mathbf{x}_i^*$ for $i = 0,1,\cdots,M-1$ are the states that the system converges to.

Lemma 1

The following statements are equivalent conditions for limit cycles:

(i) $\mathbf{s}(k_0 + M + i) = \mathbf{s}(k_0 + i) \ \ \forall i \geq 0$.

(ii) $\lim_{k \to +\infty} \mathbf{x}(kM + k_0 + i) = \mathbf{x}_i^*$ for $i = 0,1,\cdots,M-1$.

(iii) $\mathbf{x}(0) \in \Xi_1 \equiv \{\mathbf{x}(0) : \forall k \geq 0, \text{and } i = 0,1,\cdots,M-1,\ Q(\mathbf{x}(kM + k_0 + i)) = Q(\mathbf{x}_i^*)\}$.

Proof: (please see Appendix A)

Lemma 1 associates the steady state of periodic output with a specific set of initial conditions and a corresponding dynamical behavior of the system. According to Lemma 1, we can easily see that the trajectories will converge to the same set of fixed points $\{\mathbf{x}_0^*, \mathbf{x}_1^*, \cdots, \mathbf{x}_{M-1}^*\}$ after the M iterations of the map when the steady state is reached, and the periodicity of the steady states of the output sequence is equal to the number of fixed points on the phase plane. Note that fixed point is periodic point with period equal to 1. That implies that all the fixed points (more than or equal to 2) cannot be in the same quadrant. For example, if $M = 2$, then there are two fixed points on the phase plane and these two fixed points are located in different quadrants. Otherwise, if the fixed points are in the same quadrant, all the quantized system states will be the same and the periodicity of the symbolic sequences will drop.

The significance of Lemma 1 is that it provides useful information for estimating the periodicity of the steady state of output sequences via the phase portrait. We can estimate the periodicity of output sequences simply by counting the number of fixed points on the phase portrait. Moreover, Lemma 1 provides useful information to the SDM designers to operate a SDM so that a near fractal or a near chaotic behavior [77] is exhibited. If the value of M is so large that the difference of the phase portraits between the near fractal and the real fractal behaviors [77], or that between the near chaotic and the real chaotic behaviors, are visually indistinguishable, then near fractal or near chaotic behavior [77] is exhibited.

It is worth noting that although the system state is converging to a periodic orbit, it never reaches these periodic points unless the system state started on the periodic orbit in the first place. That means, the system state is aperiodic even though the output sequence is eventually periodic. This result is different from the case when $r = 1$ and θ is a rational multiple of π. For that case, the system state is periodic.

Moreover, according to (3.7) and (3.8), $\mathbf{x}_i^*$ for $i = 0,1,\cdots,M-1$ depends directly on $\mathbf{s}(i)$. It depends on $\mathbf{x}(0)$ indirectly via $\mathbf{s}(i)$ for $i = 0,1,\cdots,M-1$.

When $M = 1$, the output sequence will become constant and there is only a single fixed point on the phase portrait. The trajectory will converge to this fixed point, denoted as $\mathbf{x}^*$. From (3.7), since $M = 1$, we have:

$$\begin{aligned}\mathbf{x}^* &= \mathbf{T}\left(\lim_{p\to+\infty}\sum_{m=0}^{p-1}\mathbf{D}^m\right)\mathbf{T}^{-1}\mathbf{B}\left(\mathbf{u}-\mathbf{s}(k_0)\right)\\ &= \left(\lim_{p\to+\infty}\sum_{m=0}^{p-1}\mathbf{A}^m\right)\mathbf{B}\left(\mathbf{u}-\mathbf{s}(k_0)\right)\\ &= (\mathbf{I}-\mathbf{A})^{-1}\left(\mathbf{I}-\lim_{p\to+\infty}\mathbf{A}^p\right)\mathbf{B}\left(\mathbf{u}-\mathbf{s}(k_0)\right)\\ &= (\mathbf{I}-\mathbf{A})^{-1}\mathbf{B}\left(\mathbf{u}-\mathbf{s}(k_0)\right)\end{aligned},$$

where $\mathbf{I}$ is a 2×2 identity matrix. There is an affine linear relationship between the quantizer input and the input step size. That is, $\mathbf{x}^*$ is affine linear with respect to $\mathbf{u}$.

For the corresponding linear system, if the fixed point behavior occurs, then

$$\hat{\mathbf{x}}^* = \mathbf{A}\hat{\mathbf{x}}^* + \mathbf{B}\mathbf{u},$$

which implies that the system states will converge to $(\mathbf{I}-\mathbf{A})^{-1}\mathbf{B}\mathbf{u}$. Note that $\hat{\mathbf{x}}^* \neq \mathbf{x}^*$. Comparing these two values, there are DC shifts, which is known as affine linear, and the DC shifts are exactly dropped at the output sequences, that is:

$$\hat{\mathbf{x}}^* - \mathbf{x}^* = (\mathbf{I}-\mathbf{A})^{-1}\mathbf{B}\mathbf{u} - (\mathbf{I}-\mathbf{A})^{-1}\mathbf{B}\left(\mathbf{u}-\mathbf{s}(k_0)\right) = (\mathbf{I}-\mathbf{A})^{-1}\mathbf{B}\mathbf{s}_{k_0}, \qquad (3.9)$$

in which

$$\mathbf{s}(k) = \mathbf{s}_{k_0} \text{ for } k \geq k_0. \qquad (3.10)$$

In addition, this phenomenon is quite different for lowpass SDMs. For a stable lowpass SDM, the average value of the output sequence will approximate that of the input signal even though limit cycle behavior occurs. On the other hand, for a stable bandpass SDM, the average value of the output sequence does not have a simple relationship with that of the input signal. This is because some poles of the lowpass SDM are located at $z=1$. In order for a lowpass SDM to be stable, the average DC value of the input of the loop filter has to be zero. Otherwise, resonance will occur. This implies that the average DC value of the input has to be equal to that of the output sequence. On the other hand, for the bandpass SDM, since the poles are located at the natural frequency of the loop filter, in order for the bandpass SDM to be stable, the AC component (the component located at the natural frequency of the loop filter) of the input of the loop filter has to be zero. Otherwise, resonance will occur. This implies that the AC component of the input has to be equal to that of the output sequence. However, there is

no simple relationship between the DC component of the input of the bandpass SDM and that of the output sequence.

Figure 3.2 shows the plot of the average DC output value versus the average DC input value with $r = 0.99$, $\theta = \cos^{-1}(-0.158532)$ and $\mathbf{x}(0) = [0 \quad 0]^T$, where the input is a step signal.

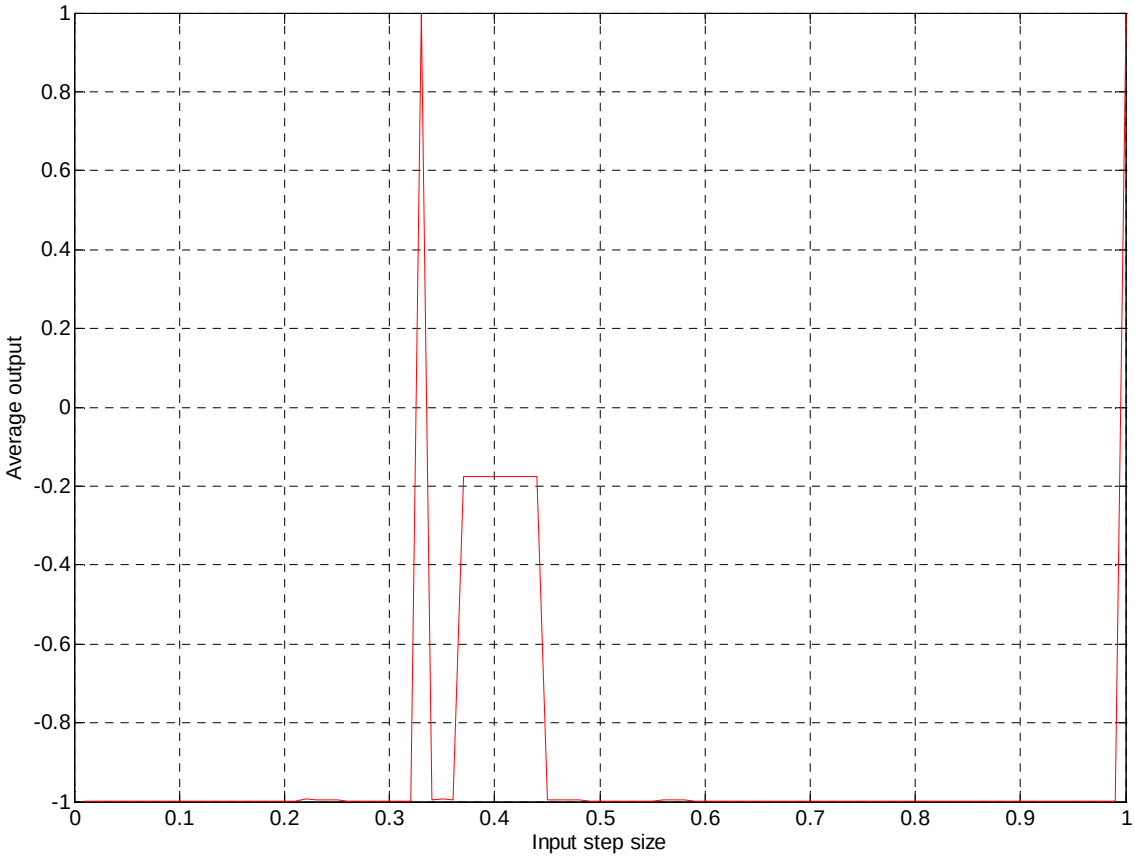


Figure 3.2 Plot of average DC output value versus average DC input value.

According to Figure 3.2, the output bitstream exhibits a fixed point behavior for the input step size outside the range between 0.36 and 0.44. This is because for fixed point behavior, the average value of the output sequence is equal to the sign of the fixed point, which is either 1 or -1.

For the input step size between 0.36 and 0.44, the output bitstream appears to be a limit cycle. For the limit cycle behavior, the value of the output sequence at some time indices is equal to 1 and that of the remaining time indices is equal to -1. Although the average value of the output sequence is also equal to the average value of the sign of the

periodic state variables, the average value of the output sequence is strictly between 1 and -1.

If we change the initial condition, then the plot of Figure 3.2 will be very different because the range of the input step size corresponding to limit cycle behavior is different for different initial conditions. Here, we need Lemma 1 to characterize the set of initial conditions corresponding to limit cycle behavior.

Although the nonlinearity is always activated, the dependence of r on the rate of convergence becomes evident when the output sequence becomes steady. This is because the DC term does not affect the rate of convergence. However, if we examine the transient response of the system, that is, the time duration before the output sequence become steady, the system dynamics can be very complex. An example in Figure 3.3 shows the response of the state variables of a bandpass SDM with

$$r = 0.9999,\ \theta = \cos^{-1}(-0.158532),\ \mathbf{u} = -0.3[1 \quad 1]^T \text{ and } \mathbf{x}(0) = [0 \quad 0.5]^T. \quad (3.11)$$

The state variable is $\mathbf{x}_1(k)$ is converging to a fixed value and the output sequence will become constant for $k \geq 2154$.

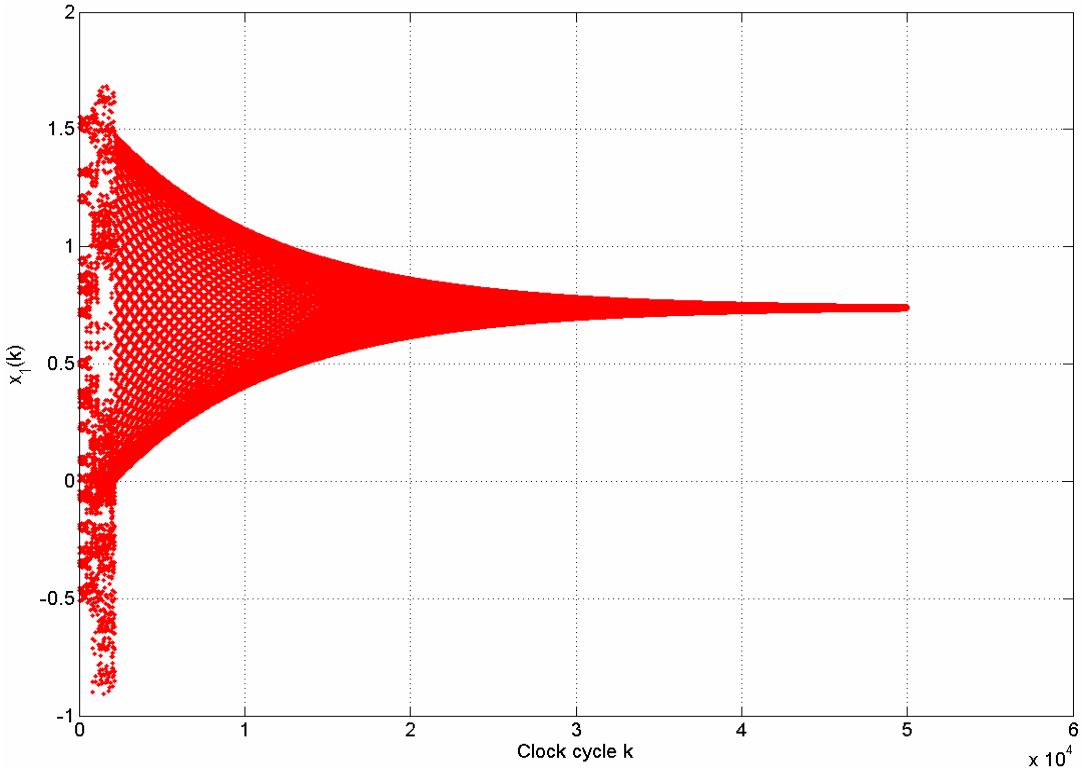


Figure 3.3 The state variable $x_1(k)$.

Figure 3.4 shows the state trajectory of another bandpass SDM with

$$r = 0.99,\ \theta = \cos^{-1}(-0.158532),\ \mathbf{u} = -0.3[1 \quad 1]^T \text{ and } \mathbf{x}(0) = [0 \quad 0.5]^T. \qquad (3.12)$$

The state trajectory is converging to two fixed points and the output sequence become periodic with period 2 for $k \geq 3$.

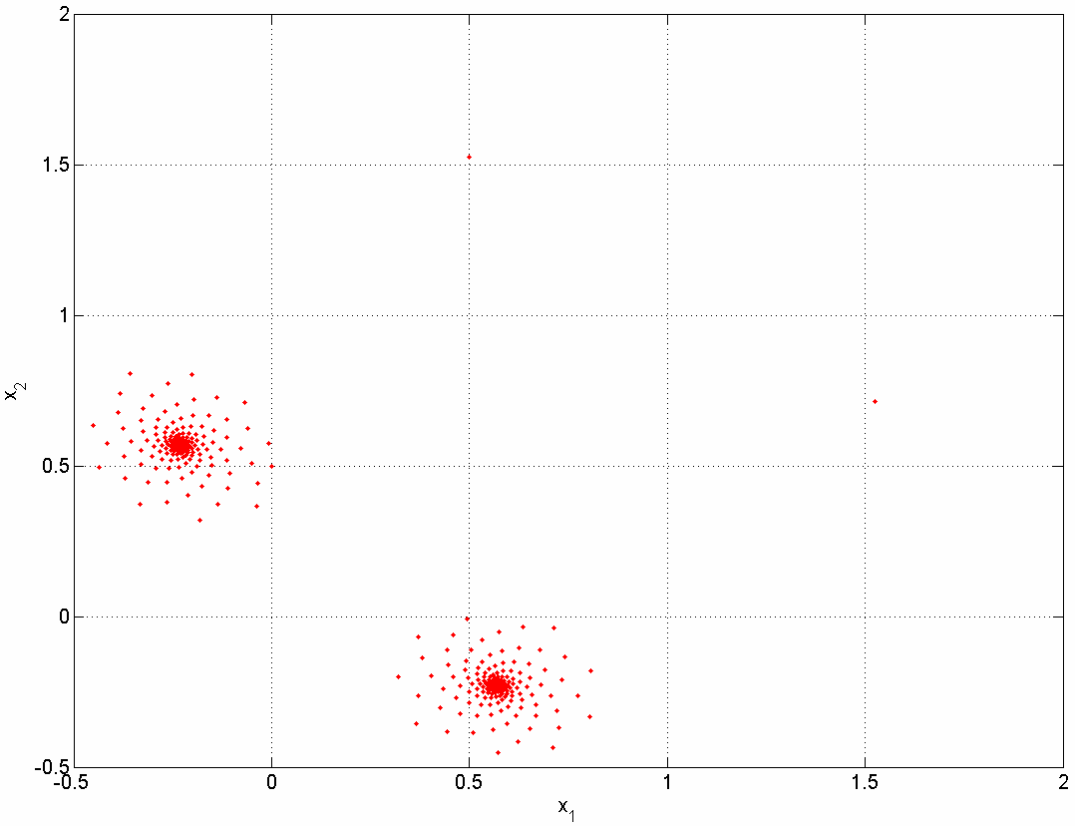


Figure 3.4 The phase portrait when $M = 2$.

Although Lemma 1 gives the necessary and sufficient conditions for the occurrence of limit cycles, it is not easy to check whether a periodic sequence is admissible or not. To address this issue, define

$$\mathbf{s} \equiv \left[\mathbf{s}(k_0)^T \quad \mathbf{s}(k_0+1)^T \quad \cdots \quad \mathbf{s}(k_0+M-1)^T\right]^T, \qquad (3.13)$$

and

$$\mathbf{K} \equiv \begin{bmatrix} \mathbf{A}^{M-1} & \cdots & \mathbf{A} & \mathbf{I} \\ \mathbf{I} & \ddots & & \mathbf{A} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{A}^{M-2} & \cdots & \mathbf{I} & \mathbf{A}^{M-1} \end{bmatrix}. \qquad (3.14)$$

Lemma 2

If a periodic sequence is admissible, then

$$Q\left(\left[\left(\left(\mathbf{I}-\mathbf{A}\right)^{-1}\mathbf{Bu}\right)^{T} \quad \cdots \quad \left(\left(\mathbf{I}-\mathbf{A}\right)^{-1}\mathbf{Bu}\right)^{T}\right]^{T}-\mathbf{K}\left(diag\left\{\left(\mathbf{I}-\mathbf{A}^{M}\right)^{-1}\mathbf{B},\cdots\left(\mathbf{I}-\mathbf{A}^{M}\right)^{-1}\mathbf{B}\right\}\mathbf{s}\right)\right)=\mathbf{s}. \quad (3.15)$$

where Q is the quantization function.

Proof: (please see Appendix B)

The importance of Lemma 2 is that it provides information to check whether a periodic sequence is admissible or not. In other words, Lemma 2 is useful to check whether a given limit cycle occurs or not for a given set of filter parameters. For a given symbolic sequence, if there exists initial condition such that the output of the quantizer is exactly equal to the symbolic sequence, then the symbolic sequence is admissible.

Note that the relationship between the occurrence of a limit cycle, and the corresponding initial condition $\mathbf{x}(0)$, input step size $\mathbf{u}$ and the filter parameters r and θ, is governed by Lemma 1 and Lemma 2. That means if $\mathbf{u}$, r, θ and $\mathbf{x}(0)$ satisfy Lemma 1 and Lemma 2, then the limit cycle will occur. On the other side, if any of these values do not satisfy Lemma 1 or Lemma 2, then that limit cycle will not occur. The occurrence of the limit cycle depends on the center frequency and the bandwidth of the loop filter via parameters r and θ in $\mathbf{A}$ and $\mathbf{B}$. θ is the center frequency and the bandwidth of the filter is uniquely determined by r and θ as follows. The bandwidth of the filter is defined as the range of the frequency that the magnitude response of the filter being larger than half of its peak value. For this loop filter, the maximum gain of the filter is

$$\left|\frac{2r\cos\theta e^{-j\theta}+r^{2}e^{-2j\theta}}{1-2r\cos\theta e^{-j\theta}+r^{2}e^{-2j\theta}}\right|.$$

Suppose that

$$\left|\frac{2r\cos\theta e^{-j(\theta+\Delta\theta)}+r^{2}e^{-2j(\theta+\Delta\theta)}}{1-2r\cos\theta e^{-j(\theta+\Delta\theta)}+r^{2}e^{-2j(\theta+\Delta\theta)}}\right|=\frac{1}{2}\left|\frac{2r\cos\theta e^{-j\theta}+r^{2}e^{-2j\theta}}{1-2r\cos\theta e^{-j\theta}+r^{2}e^{-2j\theta}}\right|,$$

then the bandwidth is $2\Delta\theta$.

In summary, the importance of Lemma 1 and Lemma 2 is for the estimation of the period of a given limit cycle. Firstly, the relationship between the occurrence of the given limit cycle and the dynamical behavior of the system is characterized. We reveal that the occurrence of a known limit cycle with period M is equivalent to the state trajectory

converging to M fixed points. Secondly, the condition we derived is a necessary and sufficient condition. Thirdly, we do not assume that the set of initial conditions generating that limit cycle is convex. Indeed, they are usually not convex. Lemma 1 requires the initial condition, which is usually unknown, for the estimation of the occurrence of a known limit cycle. Actually, it is impossible or it may be too complicated to determine the range of initial conditions that leads to a known limit cycle. This is because the set of initial conditions generating a known limit cycle is not convex. Indeed, the set of initial conditions generating a known limit cycle consists of many disjoint sets of initial conditions. Hence, there are many ranges of initial conditions generating a known limit cycle. However, this problem does not occur when we also apply Lemma 2, because Lemma 2 does not require the initial condition for the estimation of the occurrence of a known limit cycle.

Note that we can find all possible limit cycles for a bandpass SDM via Lemma 1 and Lemma 2. From Lemma 1 and Lemma 2, we can see that the occurrence of a given limit cycle is nonlinearly dependent on the input step size, so the relationship between the number of limit cycles with period M and the input step size should not be linear. In contrast to the lowpass SDM, it was found that the relationship between the number of limit cycles with period M and the input step size is approximately linear.

Figure 3.5 shows the plot of the number of different symbolic sequences against the period of the symbolic sequences for a bandpass SDM with $r = 1$, $\theta = \cos^{-1}(-0.158532)$ and $\mathbf{u} = -0.3[1 \quad 1]^T$. It can be seen from Figure 4.5 that most of the symbolic sequences do not have more than one type of limit cycle. This means in almost all cases we can easily predict whether a known limit cycle will occur or not. Moreover, this is also important to be noticed when we apply the fuzzy impulsive control technique (Chapter IV). It is because if there are more than one type of limit cycle, then the number of possible states that the impulse can reset to, so that the limit cycle can be destroyed, will be greatly reduced.

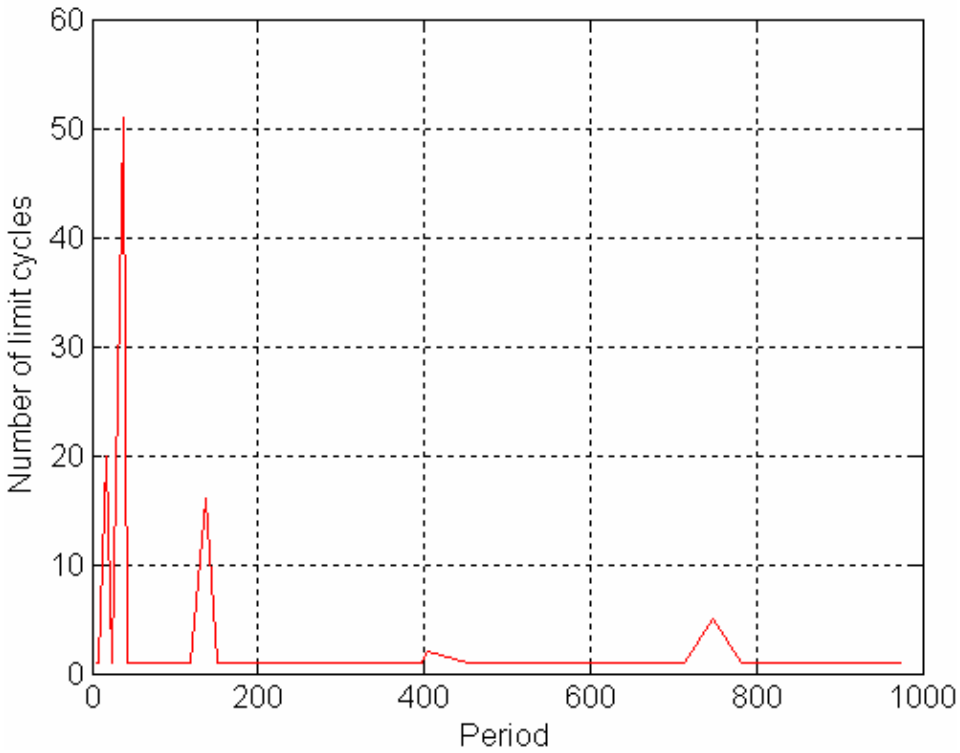


Figure 3.5 Plots of the number of different symbolic sequences against the period of the symbolic sequences.

It is found that the value of the period that gives rise to a large number of limit cycles is 38. For limit cycles with long period, we can extract the period of the limit cycle from the output spectrum using fast Fourier transform (FFT). Although this is just an approximated method and cannot guarantee that all limit cycles are found, we can, at least, find all limit cycles with period up to 25.

With Lemma 2 and Figure 3.5, we can confirm the occurrence of some limit cycles. For example, putting $s=1$ and $s=-1$, and $M=1$, for a bandpass SDM with $r=0.999999$, $\theta=\cos^{-1}(-0.158532)$ and $u=-0.3$, Lemma 2 is satisfied for $s=1$, or $s=-1$, where the period, $M=2$. Putting $s=[1 \quad -1]$ and $s=[-1 \quad 1]$, and $M=2$, for the same bandpass SDM, Lemma 2 is satisfied for $s=[1 \quad -1]$ and $s=[-1 \quad 1]$, where the period, $M=2$. Similar checking procedures can also be performed to check limit cycles with period 7 and period 11.

3.3 Near fractal or near chaotic behaviors

By substituting $\mathbf{A}$, $\mathbf{B}$, $\mathbf{u}$, $\mathbf{K}$ and $\mathbf{s}$ into equation (3.15), we have:

$$Q\left(\begin{array}{l}\dfrac{(2\cos\theta - r)ru}{1-2r\cos\theta + r^2}\begin{bmatrix}1\\1\end{bmatrix} - \dfrac{1}{\sin\theta\left(1-2r^M\cos M\theta + r^{2M}\right)}\cdot \\ \displaystyle\sum_{j=0}^{M-1}\begin{bmatrix}s_1(k_0+\mathrm{mod}(i+j,M))\\ s_2(k_0+\mathrm{mod}(i+j,M))\end{bmatrix}\cdot \\ \displaystyle\sum_{j=0}^{M-1}\begin{bmatrix}-r^{M-j}\left(r^M\sin(j+1)\theta+\sin(M-j-1)\theta\right) & 2r^{M-1-j}\cos\theta\left(r^M\sin(j+1)\theta+\sin(M-j-1)\theta\right)\\ -r^{M+1-j}\left(r^M\sin j\theta+\sin(M-j)\theta\right) & 2r^{M-j}\cos\theta\left(r^M\sin j\theta+\sin(M-j)\theta\right)\end{bmatrix}\end{array}\right)$$

$$=\begin{bmatrix}s_1(k_0+\mathrm{mod}(i+j,M))\\ s_2(k_0+\mathrm{mod}(i+j,M))\end{bmatrix}, \tag{3.16}$$

for $i = 0,1,\cdots,M-1$. It is obvious that limit cycles with longer periods are more difficult to exist or occur with good stability. However, since $\mathbf{s}(k) \in \left\{[1 \quad 1]^T, [1 \quad -1]^T, [-1 \quad 1]^T, [-1 \quad -1]^T\right\}$ for $k \geq 0$, from equation (3.16) we can check that the smaller value of M, the more difficult for the corresponding limit cycle to occur because equation (3.25), which is obtained from Lemma 2, cannot be satisfied for smaller value of M. This allows us to design a bandpass SDM with limit cycles which are with much longer periods, such as $M > 20$, so that we can consider the SDM exhibits near fractal or near chaotic behaviors [77] at the steady state for all the initial conditions.

Figures 3.6a-3.6c show the state trajectories of a bandpass SDM with filter parameters $r = 1-10^{-6}$ and $\theta = \cos^{-1}(-0.158532)$, input step size $\mathbf{u} = -0.3[1 \quad 1]^T$ and initial conditions $\mathbf{x}(0) = [0 \quad 0.5]^T$, $\mathbf{x}(0) = [0 \quad 0]^T$ and $\mathbf{x}(0) = [1 \quad 0]^T$, respectively. It can be seen from the figures that near fractal patterns are exhibited on the phase plane. In this case, the number of fixed points is very large that the difference of the phase portraits of the near fractal and the real fractal behaviors [77] are visually indistinguishable. Measurements of the fractal dimension are estimated as 1.78 in box counting dimension, 1.75 in information dimension, and 1.72 in correlation dimension for all these three initial conditions.

Figures 3.6d-3.6f show the state trajectories of a bandpass SDM with filter parameters $r = 0.9999$ and $\theta = 0.01$, input step size $\mathbf{u} = \dfrac{\pi}{10}[1 \quad 1]^T$ and initial conditions

$\mathbf{x}(0) = [0 \quad 0]^T$, $\mathbf{x}(0) = [1 \quad 0]^T$ and $\mathbf{x}(0) = [0 \quad 2]^T$, respectively. It can be seen from the figures that the SDM exhibits near chaotic patterns on the phase plane. In this case, the phase portraits of the near chaotic and the real chaotic behaviors are also visually indistinguishable. Comparing to the fixed point case shown in Figure 3.3, the value of r is the same.

Figures 3.6g-3.6i show the state trajectories of a bandpass SDM with filter parameters $r = 0.99$ and $\theta = 0.0001$, input step size $\mathbf{u} = 0.4[1 \quad 1]^T$ and initial conditions $\mathbf{x}(0) = [0 \quad 0]^T$, $\mathbf{x}(0) = [1 \quad 0]^T$ and $\mathbf{x}(0) = [0 \quad 1]^T$, respectively. In this case, the phase portraits of the near chaotic and the real chaotic behaviors are visually indistinguishable.

Without Lemma 2 and equation (3.16), it is not trivial to determine a set of system parameters for the design of a strictly stable bandpass SDM that can operate normally and will not end up with limit cycle behaviors. Referring to the bandpass SDM with state trajectory shown in Figure 3.4, there is a limit cycle with period equal to 2. Now we can design some other strictly stable bandpass SDMs with the same filter parameter, $r = 0.99$, which is close but still inside the unit circle, so that no limit cycle occurs (Figure 3.6).

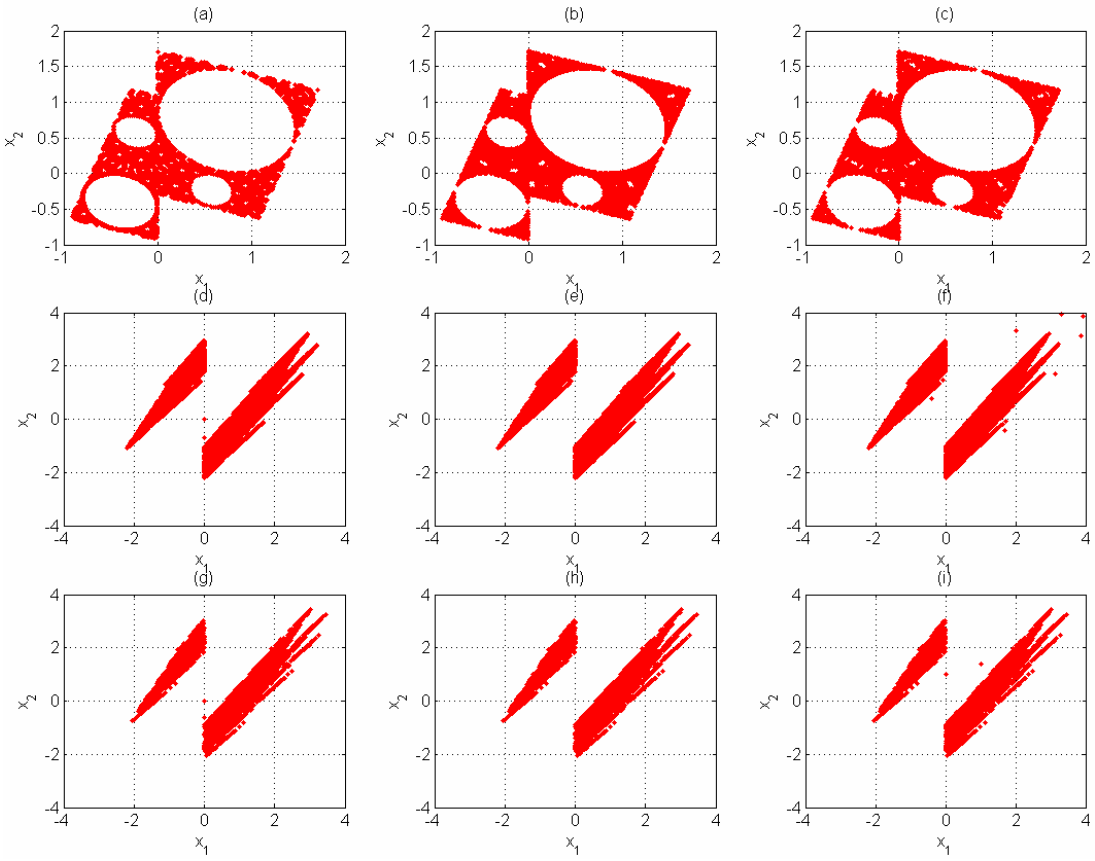


Figure 3.6 The phase portraits when the difference of the phase portraits between the near fractal and the real fractal, or the near chaotic and the real chaotic behaviors, are visually indistinguishable.

Figure 3.7 shows the spectra of the corresponding output sequences of the above examples. Unwanted audible tones in a signal mean that the signal consists of few frequency components in the frequency spectra. Therefore, when the spectra of the output sequences consist of a lot of discrete frequencies, the effect of the unwanted audible tones will be suppressed. It can be seen from the figures that the spectra of the output sequences consist of a lot of discrete frequencies, in which they are visually indistinguishable from the continuous spectra. This implies that the effect of the unwanted audible tones can be suppressed effectively.

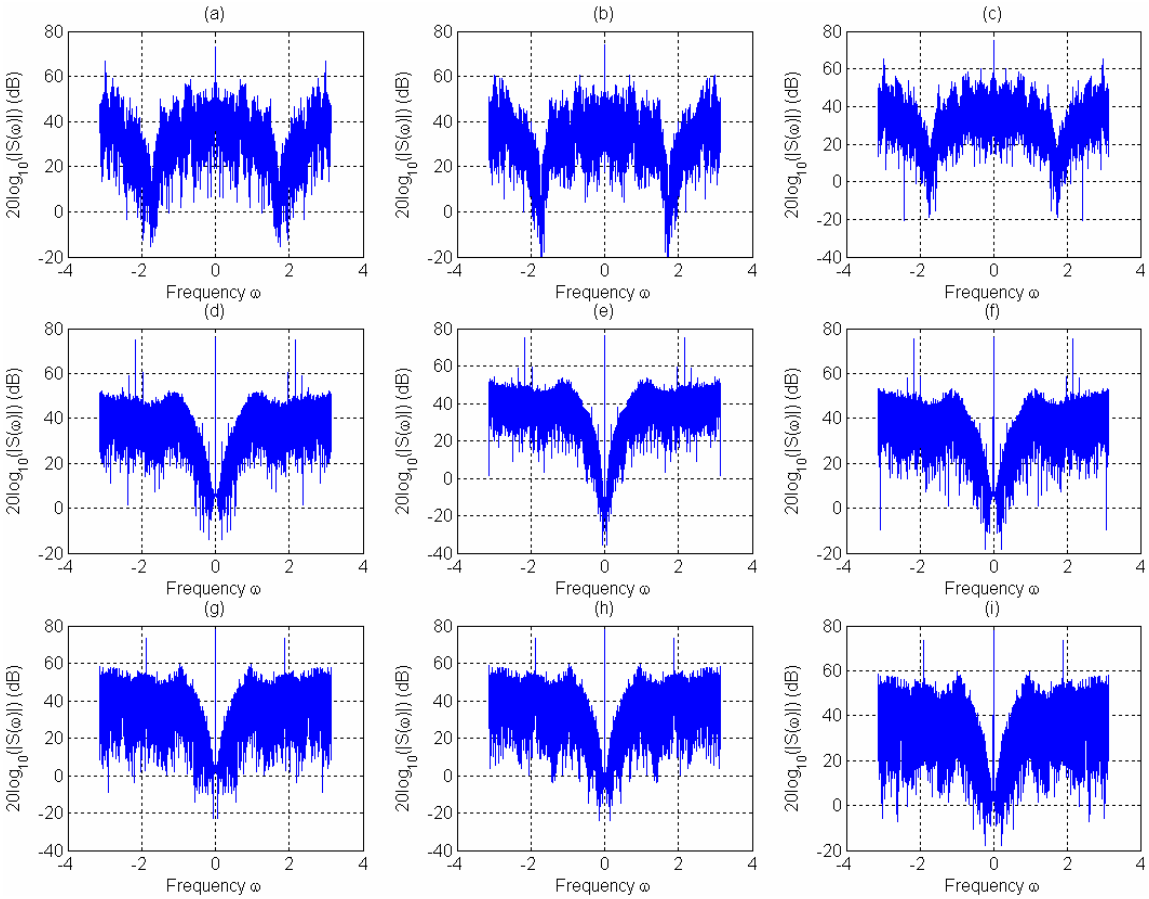


Figure 3.7 The corresponding frequency spectra of the output sequences.

Although there are some spikes in some of the spectra in Figure 3.7, by picking up the AC frequencies that produce spikes to form a set, say $\wp$, and defining the tonal suppressing ratio as the ratio of the energy of output sequences, excluding these spike components, to that of these spike components as

$$\mathrm{TSR} \equiv 10\log_{10}\left(\frac{\int_{\omega\in[-\pi,\pi]\setminus\wp}|S(\omega)|^2 d\omega}{\int_{\omega\in\wp}|S(\omega)|^2 d\omega}\right), \tag{3.17}$$

we find that the tonal suppression ratios of the above SDMs are 15.91dB, 19.27dB, 17.76dB, 6.89dB, 6.90dB, 6.59dB, 7.75dB, 7.78dB and 7.75dB, respectively. From these figures, we can conclude that the energy of the unwanted audible tones can be neglected because the energy of the rich frequency components is much stronger. For $\mathrm{TSR} = 6.59\mathrm{dB}$ and 19.27dB, the energy of the rich frequency components are 4.5599

and 84.5999 times greater than that of the corresponding unwanted audible tones, respectively.

In addition, since all the simulations are carried out using MATLAB under a 64 bit computer, the numerical rounding error is insignificant in reference to the distance between the poles and the unit circle. For example, the numerical error due to a 64 bit computer is 2^{-64}, while the distance between the pole and the unit circle is 10^{-4} for the similar strictly stable bandpass SDM with $r = 0.9999$. We can see that the ratio is just 5.42×10^{-16}. The argument suggests that rounding errors might be of minor importance.

3.4 Conclusions

One possible implication of the results obtained in this research is that it is not necessary to place unstable poles to the bandpass SDM to generate signal with rich frequency spectrum in order to suppress the unwanted tones from quantizer. We have shown that near fractal or near chaotic signal can be generated via strictly stable poles. Since the spectra of the output sequence consists of a lot of frequency components which are visually indistinguishable from the continuous spectrum, we have shown that the unwanted tones can be suppressed effectively by these rich frequency components and bounded states of the SDM can be guaranteed.

CHAPTER IV. FUZZY IMPULSIVE CONTROL OF HIGH ORDER SDMS

As discussed in Section 1.3.1, limit cycles are more prevalent with low order filters, while unbounded system states may be observed when the inputs of the SDMs are overloaded and the order of the loop filter is high. In order to solve these problems, control is required. However, as discussed in Section 1.4, most common and traditional control techniques fail to stabilize SDMs or result in the occurrence of limit cycles when the input is increased.

4.1 Definition and advantages of impulsive and fuzzy controls

The purpose of an impulsive control is to reset the system state to somewhere position in the state space determined by some control laws. Whereas the clipping method, the system states are always reset to certain fixed values. In the traditional control strategies, the control force is added to the input signal and uses the input signal to influence the system states. Therefore, the mechanism of the impulsive control strategy is different from the traditional ones.

As there are usually infinite system states in the state space, fuzzy rules are employed to determine those system states that the impulsive controller should be reset to. This control technique is very powerful because conventional control strategies require precise information of the system. However, system information sometimes cannot be obtained precisely. For example, if the temperature is hot, then the power of the air conditioner should increase. But the word “hot” is very fuzzy. Different people have different interpretation of the hotness. Fuzzy rules are those rules formulated based on the expert and heuristic knowledge.

Since the SDM consists of a quantizer, nonlinear behaviors, such as fractal and chaotic behaviors, may occur. With the practical consideration on the boundedness of the system states and a heuristic measure on the strength of audible clicks, it is very difficult to determine the suitable system states analytically. To solve this problem, a fuzzy approach is employed to simplify the complicated problems and capture the heuristic knowledge in the system. The main advantages of fuzzy impulsive control are to avoid

the occurrence of limit cycle behaviors and minimize the effect of audible clicks with the guarantee of bounded system states. Moreover, if an invariant set exists, the impulsive control will be very efficient if our control goal is only to maintain the boundedness of the system states. This is because the control force applied at single time instant is sufficient to reset the system states to maintain the boundedness of the system states within the invariant set forever as long as the input signal does not change. On the other hand, clipping actions are required continuously in order to maintain the boundedness of the system states.

4.2 Analysis on limit cycle behaviors and boundedness of system states

Consider an interpolative SDM shown in Figure 1.3. Here, $F(z)$ is assumed to be real, causal, rational, proper and with the order of the polynomial of z^{-1} in the numerator being equal to that in the denominator as well as there is a delay multiplied in the numerator. We make those assumptions because this type of SDMs is commonly used in industry [4]. Denote the coefficients in the denominator and numerator of $F(z)$ as, respectively, a_i for $i = 0,1,\cdots,N$ and b_j for $j = 1,\cdots,N$, where N is the order of the loop filter. Then

$$F(z) = \frac{\sum_{j=1}^{N} b_j z^{-j}}{\sum_{i=0}^{N} a_i z^{-i}}. \tag{4.1}$$

Since we are based on the feedforward structure of the SDM, without loss of generality, we assume that the loop filter is realized via the direct form realization. For the other minimal realizations, they can be converted to the direct form realization using simple transformations. Using a similar approach as discussed in Section 3.1, we have:

$$F(z) = \frac{Y(z)}{U(z) - S(z)} = \frac{\sum_{j=1}^{N} b_j z^{-j}}{\sum_{i=0}^{N} a_i z^{-i}},$$

which implies that

$$Y(z)\sum_{i=0}^{N} a_i z^{-i} = (U(z) - S(z))\sum_{j=1}^{N} b_j z^{-j},$$

where $Y(z)$, $U(z)$ and $S(z)$ are the z-transforms of the output of the loop filter, the input signal and the output of the quantizer, respectively. By expressing the equation in the form of a difference equation, we have:

$$\sum_{i=0}^{N} a_i y(k-i) = \sum_{j=1}^{N} b_j (u(n-j) - s(n-j))$$

or

$$y(k) = \frac{1}{a_0}\left(\sum_{j=1}^{N} b_j (u(n-j) - s(n-j)) - \sum_{i=1}^{N} a_i y(k-i)\right).$$

By expressing the equation in the matrix form, we have:

$$\begin{bmatrix} y(k-N+1) \\ \vdots \\ y(k) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \\ -\frac{a_N}{a_0} & \cdots & \cdots & \cdots & -\frac{a_1}{a_0} \end{bmatrix} \begin{bmatrix} y(k-N) \\ \vdots \\ y(k-1) \end{bmatrix} +$$

$$\begin{bmatrix} 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 \\ \frac{b_N}{a_0} & \cdots & \cdots & \cdots & \frac{b_1}{a_0} \end{bmatrix} \left(\begin{bmatrix} u(k-N) \\ \vdots \\ u(k-1) \end{bmatrix} - \begin{bmatrix} s(k-N) \\ \vdots \\ s(k-1) \end{bmatrix} \right).$$

By letting

$$\mathbf{x}(k) \equiv [x_1(k), \quad \cdots, \quad x_N(k)]^T \equiv [y(k-N), \quad \cdots, \quad y(k-1)]^T \tag{4.2}$$

as the system state of the SDM,

$$\mathbf{u}(k) \equiv [u(k-N), \quad \cdots, \quad u(k-1)]^T, \tag{4.3}$$

$$\mathbf{s}(k) \equiv [s_1(k), \quad \cdots, \quad s_N(k)]^T \equiv [Q(y(k-N)), \quad \cdots, \quad Q(y(k-1))]^T, \tag{4.4}$$

$$\mathbf{A} \equiv \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \\ -\dfrac{a_N}{a_0} & \cdots & \cdots & \cdots & -\dfrac{a_1}{a_0} \end{bmatrix} \tag{4.5}$$

and

$$\mathbf{B} \equiv \begin{bmatrix} 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 \\ \dfrac{b_N}{a_0} & \cdots & \cdots & \cdots & \dfrac{b_1}{a_0} \end{bmatrix}, \tag{4.6}$$

in which Q is a one-bit quantizer defined as follows,

$$Q(y) \equiv \begin{cases} 1 & y \geq 0 \\ -1 & \text{otherwise} \end{cases}, \tag{4.7}$$

then the SDM can be described by the following state space equation:

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{B}(\mathbf{u}(k) - \mathbf{s}(k)) \text{ for } k \geq 0. \tag{4.8}$$

Since the OSR of the SDM is usually very high, the input can be approximated as a step signal. Hence, we further assume that $\mathbf{u}(k) = \bar{\mathbf{u}}$ for $k \geq 0$.

In many practical situations, the magnitude of the state variables of the SDM should not be larger than certain values, otherwise the devices will be damaged. For the direct form realization, since all the state variables are the delay versions of the output of the loop filter, we denote the bounds of the state variables as V_{cc}. That is, $|x_i(k)| < V_{cc}$ for $i = 1,2,\cdots,N$ and $k \geq 0$. Otherwise, the SDM may yield an unwanted behavior. Denote B_o as the set of allowable system states. That is, $B_o = \{\mathbf{x} : |x_i| < V_{cc} \text{ for } i = 1,2,\cdots,N\}$.

For most applications, such as audio applications [4], it is undesirable to have periodic output sequences. Hence, before we propose the fuzzy impulsive control strategy, the conditions for exhibiting limit cycle behavior and the corresponding set of initial conditions will be discussed in Lemma 3. This is essential for formulating a fuzzy membership function which can avoid the occurrence of limit cycle behavior.

Since $a_0 \neq 0$ and $a_N \neq 0$, otherwise the order of the filter will be dropped, the columns of matrix $\mathbf{A}$ are linearly independent and the eigen decomposition of matrix $\mathbf{A}$ exists. There also exists a full rank matrix $\mathbf{T}$ and a diagonal matrix $\mathbf{D}$ which consist of the eigenvectors and eigenvalues of matrix $\mathbf{A}$, respectively, such that $\mathbf{A} = \mathbf{TDT}^{-1}$. Let λ_i and ξ_i for $i = 1,2,\cdots,N$ be the eigenvalues and the corresponding eigenvectors of matrix $\mathbf{A}$. Let n_d be the number of eigenvalues of matrix $\mathbf{A}$ on the unit circle, where their phases are integer multiples of $\frac{2\pi}{P}$, that is, $\lambda_{i+N-n_d} = e^{\frac{j2\pi k_i}{P}}$ for $k_i \in Z$ and $i = 1,2,\cdots,n_d$. Let L_i for $i = 1,2,\cdots,N$, be the i^{th} row of

$$\sum_{j=0}^{P-1} \mathbf{A}^{P-1-j}\mathbf{B}\left(\mathbf{u}\left(k_0 + j\right) - \mathbf{s}\left(k_0 + j\right)\right), \tag{4.9}$$

where $P \in Z^+$ and $k_0 \geq 0$. Let $\mathbf{r}_j$ for $j = 1,2,\cdots,N$ be the j^{th} row of $\mathbf{I} - \mathbf{A}^P$, where $\mathbf{I}$ is an $N \times N$ identity matrix. Denote

$$\Psi_P \equiv \left\{\mathbf{x}(0) : \mathbf{r}_i \mathbf{x}(k_0) = L_i \text{ for } i = 1,2,\cdots,N - n_d\right\}. \tag{4.10}$$

Lemma 3

Suppose there are $N - n_d$ linearly independent rows in the matrix $\mathbf{I} - \mathbf{A}^P$ and these $N - n_d$ linearly independent rows are the first $N - n_d$ rows of the matrix $\mathbf{I} - \mathbf{A}^P$, that is, an $\exists c_{i,n} \in \Re$ for $i = 1,2,\cdots,N - n_d$ and $n = 1,2,\cdots,n_d$ such that $\sum_{i=1}^{N-n_d} c_{i,n}\mathbf{r}_i = \mathbf{r}_{N-n_d+n}$. If $\Psi_P \neq \emptyset$ and $\sum_{i=1}^{N-n_d} c_{i,n}L_i = L_{N-n_d+n}$ for $n = 1,2,\cdots,n_d$, then the SDM will exhibit limit cycle behavior with period P, and Ψ_P is the corresponding nonempty set of initial conditions. If $\Psi_P = \emptyset$ or $\exists n \in \left\{1,2,\cdots,n_d\right\}$ such that $\sum_{i=1}^{N-n_d} c_{i,n}L_i \neq L_{N-n_d+n}$, then there will not exist any fixed point or periodic state sequence.

Proof: (please see Appendix C)

The importance of this lemma is to characterize the set of initial condition that corresponds to the limit cycle behaviors with period P for $k \geq k_0$. This set of initial conditions will be used for the formulation of fuzzy rules.

This result is a generalization of the existing results [45]. The existing results mainly consider the DC pole cases, that is $k_i = 0$ for $i = 1,2,\cdots,n_d$. However, we reveal that even though there is no DC pole, but if there exists some poles on the unit circle with phases that are nonzero integer multiples of $\frac{2\pi}{P}$, then the matrix $\mathbf{Q}$ will also drop rank. Besides, when there is more than one DC pole in the loop filter transfer function, if the geometric multiplicity of this pole as an eigenvalue of $\mathbf{A}$ is equal to its algebraic multiplicity, then the eigen decomposition of matrix $\mathbf{A}$ will exist and Lemma 3 will still be applied.

Define the forward and backward dynamics of the system as $\aleph_{\mathrm{f}} : \Re^N \to \Re^N$ and $\aleph_{\mathrm{b}} : \Re^N \to \Re^N$, respectively. That is:

$$\mathbf{x}(k+1) \equiv \aleph_{\mathrm{f}}(\mathbf{x}(k)) \text{ in which } \mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{B}(\bar{\mathbf{u}} - Q(\mathbf{x}(k))) \tag{4.11}$$

and

$$\mathbf{x}(k-1) \equiv \aleph_{\mathrm{b}}(\mathbf{x}(k)) \text{ in which } \mathbf{x}(k) = \mathbf{A}\mathbf{x}(k-1) + \mathbf{B}(\bar{\mathbf{u}} - Q(\mathbf{x}(k-1))), \tag{4.12}$$

respectively. Denote

$$x'(k) \equiv b_N \bar{u} + \sum_{i=1}^{N-1} b_{N-i}(\bar{u} - Q(x_i(k))) - \sum_{i=1}^{N} a_{N-i} x_i(k) \tag{4.13}$$

and

$$\hat{\mathbf{x}}(k) \equiv \left[\frac{x'(k) - Q(x'(k)a_N)b_N}{a_N}, \quad x_1(k), \quad \cdots, \quad x_{N-1}(k) \right]^T. \tag{4.14}$$

Then the system equation can be represented as the following matrix equation:

$$\begin{aligned}
&\mathbf{A}\hat{\mathbf{x}}(k) + \mathbf{B}(\bar{\mathbf{u}} - Q(\hat{\mathbf{x}}(k))) \\
&= \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \\ -\frac{a_N}{a_0} & \cdots & \cdots & \cdots & -\frac{a_1}{a_0} \end{bmatrix} \begin{bmatrix} \frac{x'(k) - Q(x'(k)a_N)b_N}{a_N} \\ x_1(k) \\ \vdots \\ x_{N-1}(k) \end{bmatrix} + \begin{bmatrix} 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 \\ \frac{b_N}{a_0} & \cdots & \cdots & \cdots & \frac{b_1}{a_0} \end{bmatrix} \left(\begin{bmatrix} \bar{u} \\ \vdots \\ \bar{u} \end{bmatrix} - Q\left(\begin{bmatrix} \frac{x'(k) - Q(x'(k)a_N)b_N}{a_N} \\ x_1(k) \\ \vdots \\ x_{N-1}(k) \end{bmatrix} \right) \right).
\end{aligned}$$

By solving for each row, we have

$$\mathbf{A}\hat{\mathbf{x}}(k)+\mathbf{B}(\bar{\mathbf{u}}-Q(\hat{\mathbf{x}}(k)))$$
$$=\begin{bmatrix} x_1(k) \\ \vdots \\ x_{N-1}(k) \\ -\dfrac{x'(k)-Q(x'(k)a_N)b_N}{a_0}-\sum_{i=1}^{N-1}\dfrac{a_{N-i}x_i(k)}{a_0}+\sum_{i=1}^{N-1}\dfrac{b_{N-i}(\bar{u}-Q(x_i(k)))}{a_0}+\dfrac{b_N\left(\bar{u}-Q\left(\dfrac{x'(k)-Q(x'(k)a_N)b_N}{a_N}\right)\right)}{a_0} \end{bmatrix}.$$

By grouping the terms, we have

$$\mathbf{A}\hat{\mathbf{x}}(k)+\mathbf{B}(\bar{\mathbf{u}}-Q(\hat{\mathbf{x}}(k)))$$
$$=\begin{bmatrix} x_1(k) \\ \vdots \\ x_{N-1}(k) \\ -\dfrac{b_N\bar{u}+\sum_{i=1}^{N-1}b_{N-i}(\bar{u}-Q(x_i(k)))-\sum_{i=1}^{N}a_{N-i}x_i(k)-Q(x'(k)a_N)b_N}{a_0}-\sum_{i=1}^{N-1}\dfrac{a_{N-i}x_i(k)}{a_0}+\sum_{i=1}^{N-1}\dfrac{b_{N-i}(\bar{u}-Q(x_i(k)))}{a_0}+\dfrac{b_N\left(\bar{u}-Q\left(\dfrac{x'(k)-Q(x'(k)a_N)b_N}{a_N}\right)\right)}{a_0} \end{bmatrix}$$

, and it eventually reduces to

$$\mathbf{A}\hat{\mathbf{x}}(k)+\mathbf{B}(\bar{\mathbf{u}}-Q(\hat{\mathbf{x}}(k)))$$
$$=\left[x_1(k),\ \cdots,\ x_{N-1}(k),\ x_N(k)+\frac{b_N}{a_0}\left(Q(x'(k)a_N)-Q\left(\frac{x'(k)-Q(x'(k)a_N)b_N}{a_N}\right)\right)\right]^T. \tag{4.15}$$

If $|x'(k)|>|b_N|$, then

$$Q(x'(k)-Q(x'(k)a_N)b_N)=Q(x'(k)). \tag{4.16}$$

Hence,

$$Q(x'(k)a_N)-Q\left(\frac{x'(k)-Q(x'(k)a_N)b_N}{a_N}\right)=Q(x'(k)a_N)-Q(x'(k)a_N)=0 \tag{4.17}$$

and

$$\mathbf{A}\hat{\mathbf{x}}(k)+\mathbf{B}(\bar{\mathbf{u}}-Q(\hat{\mathbf{x}}(k)))=[x_1(k),\ \cdots,\ x_N(k)]^T=\mathbf{x}(k). \tag{4.18}$$

If $|x'(k)|<|b_N|$, then

$$Q(x'(k)-Q(x'(k)a_N)b_N)=-Q(x'(k)a_N)Q(b_N) \tag{4.19}$$

and

$$Q(x'(k)a_N)-Q\left(\frac{x'(k)-Q(x'(k)a_N)b_N}{a_N}\right)=Q(x'(k)a_N)+Q(x'(k)a_N)Q(a_Nb_N). \tag{4.20}$$

If $Q(a_Nb_N)=-1$, then

$$Q(x'(k)a_N)-Q\left(\frac{x'(k)-Q(x'(k)a_N)b_N}{a_N}\right)=0 \tag{4.21}$$

and

$$\mathbf{A}\hat{\mathbf{x}}(k)+\mathbf{B}(\bar{\mathbf{u}}-Q(\hat{\mathbf{x}}(k)))=\mathbf{x}(k). \tag{4.22}$$

Hence, if $|x'(k)|>|b_N|$, or $|x'(k)|<|b_N|$ and $Q(a_N b_N)=-1$, then the backward dynamics of the SDMs can be defined as

$$\aleph_{\mathrm{b}}(\mathbf{x}(k))=\left[\frac{x'(k)-Q(x'(k)a_N)b_N}{a_N},\quad x_1(k),\quad \cdots,\quad x_{N-1}(k)\right]^T. \tag{4.23}$$

Suppose the above conditions for the existence of the backward dynamics are satisfied $\forall k\in Z$. Denote

$$\wp\equiv\{\mathbf{x}(0):\aleph_{\mathrm{f}}(\mathbf{x}(k))\in\wp \quad \text{for } k\geq 0, \quad \text{and} \quad \aleph_{\mathrm{b}}(\mathbf{x}(k))\in\wp \quad \text{for } k\leq 0\} \tag{4.24}$$

and a map $\Im:\wp\rightarrow\wp$ such that

$$\Im(\mathbf{x})\equiv\mathbf{A}\mathbf{x}+\mathbf{B}(\bar{\mathbf{u}}-Q(\mathbf{x})). \tag{4.25}$$

Lemma 4

If $|x'(k)|>|b_N|$, or $|x'(k)|<|b_N|$ and $Q(a_N b_N)=-1$, then $\wp$ will be an invariant set under $\Im$. That is, $\Im(\wp)\equiv\wp$. Hence, if $\exists k_0\in Z$ such that $\mathbf{x}(k_0)\in\wp$, then $\mathbf{x}(k)\in\wp$ $\forall k\in Z$.

Proof: (please see Appendix D)

It was reported in [24] that if an invariant set exists and there exists an initial condition in the invariant set, then the local boundedness of the system states will be guaranteed. However, it is worth noting that if $\exists k_0\in Z$ such that $\mathbf{x}(k_0)\in\Re^N\setminus\wp$ and $\mathbf{x}(k)\in\Re^N\setminus\wp$ $\forall k\in Z$, then $\mathbf{x}(k)$ may diverge. Hence, it is not sufficient to conclude the global boundedness of the system states of an SDM only from the existence of an invariant set.

The conditions for the existence of the invariant set have not been explored and this relationship is stated in Lemma 4. It is worth noting that the shape of the invariant set can be both regular and irregular, and dependent on both $\mathbf{A}$ and $\bar{\mathbf{u}}$. One example of a regular and an irregular invariant set is, respectively, the set of state trajectories when it exhibits the elliptic fractal patterns and the chaotic behaviors respectively.

The importance of Lemma 4 is that it provides information for formulating a fuzzy membership function to achieve local boundedness of the system states.

4.3 Proposed control strategy

Figure 4.1 shows the block diagrams of how the fuzzy impulsive controller influences the SDM. As discussed in Section 4.1, the fuzzy impulsive controller determines the controlled system states and resets the state variables of the loop filter to the controlled state variables via a reset circuit.

To determine the controlled system states, two step procedures are employed. The first step of the procedure is the training phase in which the invariant set and the set of system states that exhibits limit cycle behaviors are learnt through training. By generating a set of DC signals that is inputted to the system with different initial conditions, the system states are tested to see if they form an invariant set and exhibit a limit cycle behavior or not.

For the implementation of the controller, this can be done by the mean of digital control. Equation (4.33) determines the control state vector according to the initial condition. The fuzzy impulsive control law is formulated as an optimization problem discussed in equation (4.33). By using mathematical computer aided design tools such as Matlab, the impulsive control force can be evaluated. This part is carried out via a very fast computer. After the impulsive control force is calculated, this value is sent to the loop filter via input/ouput device of the computer. The value is used to reset the system states of the loop filter stored in the registers. The implementability only depends on the speed of the computer required for solving the optimization problem. This is because the reset action can be carried out instantaneously.

The second step of the procedure is the control phase in which the controlled system states are determined and the state variables are reset to the corresponding values. The details are discussed as below.

Training and Learning Phase:

Step 1: Initialization of state variables

Step 2: Try a set of inputs (eg. 100 different inputs) to optimize and train fuzzy control rules and control membership function.

If the control rules conditions of

1. the continuity of change of the state variables
2. the stability of the state variables
3. the avoidance of limit cycle

are satisfied, go to Step 3.

Otherwise, continue Step 2 with another inputs.

Step 3: Update the state variables as the controlled state variables. In the end, the membership function will be updated. The controlled SDM will be obtained.

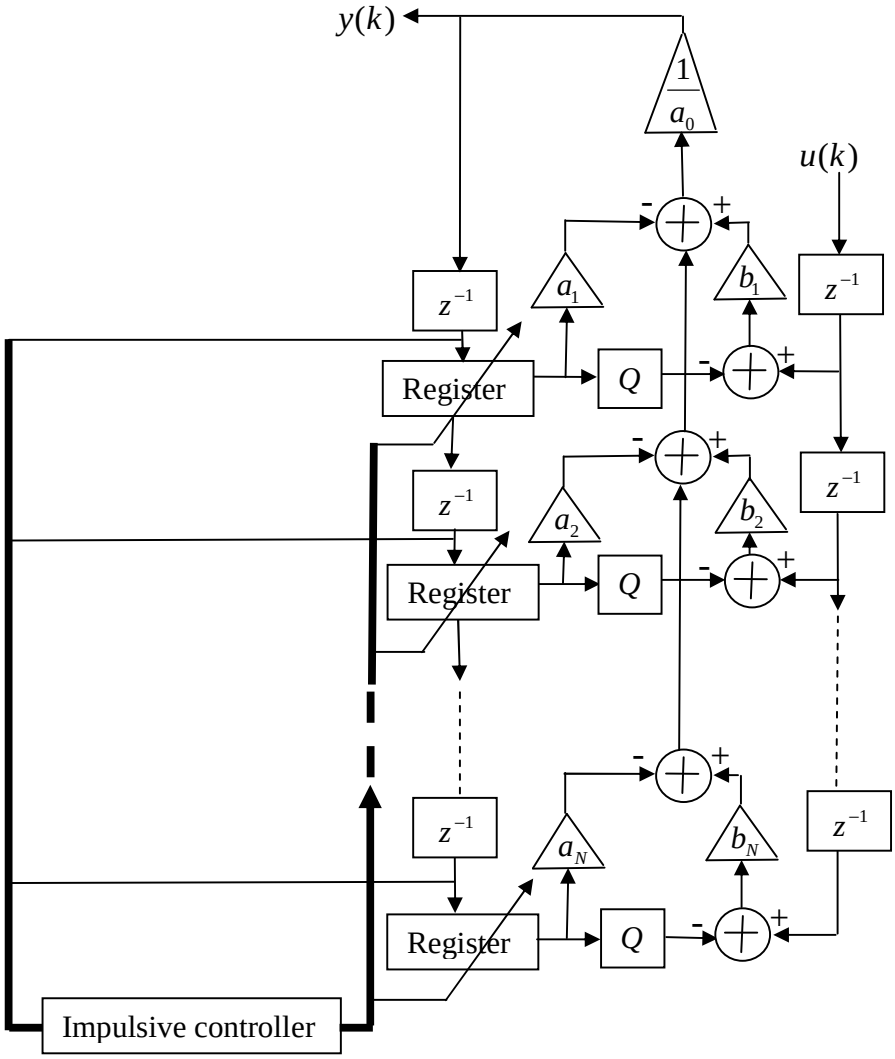


Figure 4.1 The block diagram of the interpolative SDM under the fuzzy impulsive control strategy.

For audio applications, we want to minimize the effect of audible clicks. To achieve this goal, we want to minimize the distance between the original system states $\mathbf{x}(k_0+1)$ and the controlled system states $\mathbf{x}^c(k_0+1)$. However, $\mathbf{x}(k_0+1)$ may be outside the desired bounded region B_0, so we define a vector $\mathbf{x}^r \in B_0$ such that $\left\|\mathbf{x}(k_0+1)-\mathbf{x}^r\right\|_2$ is minimum and our goal is to minimize the distance between $\mathbf{x}^c(k_0+1)$ and $\mathbf{x}^r$ via a triangular fuzzy membership function as follows:

$$\mu_{\text{distance}}\left(\mathbf{x}^c(k_0+1)\right) \equiv \left(\prod_{i=1}^{N} f_i\left(\mathbf{x}^c(k_0+1), \mathbf{x}^r\right)\right)^{\frac{1}{N}}, \tag{4.26}$$

where

$$f_i\left(\mathbf{x}^c\left(k_0+1\right),\mathbf{x}^r\right)\equiv\begin{cases}\dfrac{x_i^c\left(k_0+1\right)-V_{cc}}{x_i^r-V_{cc}} & x_i^r\le x_i^c\left(k_0+1\right)\le V_{\mathrm{cc}}\\ \dfrac{x_i^c\left(k_0+1\right)+V_{cc}}{x_i^r+V_{cc}} & -V_{\mathrm{cc}}\le x_i^c\left(k_0+1\right)\le x_i^r\\ 0 & \text{otherwise}\end{cases}, \tag{4.27}$$

Since a triangular fuzzy membership function is employed and $\mathbf{x}^r\in B_0$, $\mu_{\text{distance}}\left(\mathbf{x}^c\left(k_0+1\right)\right)=1$ when $\mathbf{x}^c\left(k_0+1\right)=\mathbf{x}^r$, $\mu_{\text{distance}}\left(\mathbf{x}^c\left(k_0+1\right)\right)=0$ when $\mathbf{x}^c\left(k_0+1\right)\in\Re^N\setminus B_0$, and $0\le\mu_{\text{distance}}\left(\mathbf{x}^c\left(k_0+1\right)\right)\le 1$ $\forall\mathbf{x}^c\left(k_0+1\right)\in B_0$. Hence, $\mu_{\text{distance}}\left(\mathbf{x}^c\left(k_0+1\right)\right)$ force the new system states $\mathbf{x}^c\left(k_0+1\right)$ to be within B_0. Note that if $\mathbf{x}\left(k_0+1\right)\in B_0$, then $\mathbf{x}^r=\mathbf{x}\left(k_0+1\right)$ and there will be no audible click effect by setting $\mathbf{x}^c\left(k_0+1\right)=\mathbf{x}^r$. Since $\mu_{\text{distance}}\left(\mathbf{x}^c\left(k_0+1\right)\right)$ captures the knowledge on the closeness between $\mathbf{x}^c\left(k_0+1\right)$ and $\mathbf{x}^r$, and the effect of audible clicks is minimized if $\mathbf{x}^c\left(k_0+1\right)$ is close to $\mathbf{x}^r$, this fuzzy membership function can minimize the effect of audible clicks.

As the local boundedness of the system states is important for many situations, according to Lemma 4, if $\left|x'(k)\right|>\left|b_N\right|$, or $\left|x'(k)\right|<\left|b_N\right|$ and $Q\left(a_N b_N\right)=-1$, then $\mathbf{x}(k)\in\wp$ $\forall k\in Z$ supposing that $\exists k_0\in Z$ such that $\mathbf{x}\left(k_0\right)\in\wp$. However, the trajectory may not be inside B_0 because $\wp$ is usually not equal to B_0. In order to guarantee that the trajectory is bounded within B_0, we want the controlled system states to be inside $\wp\cap B_0$, that is, $\mathbf{x}^c\left(k_0+1\right)\in\wp\cap B_0$. Supposing that $\wp\cap B_0\neq\emptyset$, this implies that there exists some system states such that the state responses are within the desired set of bounded system states. Denote $\mathbf{x}^p\in\wp\cap B_0$ such that $\left\|\mathbf{x}\left(k_0+1\right)-\mathbf{x}^p\right\|_2$ is minimum. For $\wp\cap B_0\neq\emptyset$, $\left|x'(k)\right|>\left|b_N\right|$, or $\left|x'(k)\right|<\left|b_N\right|$ and $Q\left(a_N b_N\right)=-1$, we define the following triangular fuzzy membership function:

$$\mu_{\text{stable}}\left(\mathbf{x}^c\left(k_0+1\right)\right)\equiv\left(\prod_{i=1}^{N} f_i\left(\mathbf{x}^c\left(k_0+1\right),\mathbf{x}^p\right)\right)^{\frac{1}{N}}, \tag{4.28}$$

Since a triangular fuzzy membership function is employed and $\mathbf{x}^p \in B_0$, $\mu_{\text{stable}}\left(\mathbf{x}^c(k_0+1)\right)=0$ when $\mathbf{x}^c(k_0+1)\in \Re^N \setminus B_0$, $\mu_{\text{stable}}\left(\mathbf{x}^c(k_0+1)\right)=1$ when $\mathbf{x}^c(k_0+1)=\mathbf{x}^p$ and $0\le \mu_{\text{stable}}\left(\mathbf{x}^c(k_0+1)\right)\le 1 \quad \forall \mathbf{x}^c(k_0+1)\in B_0$. Hence, $\mu_{\text{stable}}\left(\mathbf{x}^c(k_0+1)\right)$ forces the new system states $\mathbf{x}^c(k_0+1)$ to be within B_0. If $\mathbf{x}(k_0+1)\in \wp \cap B_0$, then $\mathbf{x}^p=\mathbf{x}(k_0+1)$. By setting $\mathbf{x}^c(k_0+1)=\mathbf{x}^p$, the criterion for the local boundedness of system states is satisfied. Since $\mu_{\text{stable}}\left(\mathbf{x}^c(k_0+1)\right)$ captures the knowledge on the closeness between $\mathbf{x}^c(k_0+1)$ and $\mathbf{x}^p$, which also reflects the closeness between $\mathbf{x}^c(k_0+1)$ and the set of system states that achieves local boundedness within the desired bounded region, this fuzzy membership function can capture the criterion for the local boundedness of the system states into the control strategy.

However, if $\wp \cap B_0 = \text{Ø}$, then $\mathbf{x}^p$ does not exist. Or if $\exists k' \in Z$ such that $|x'(k)| < |b_N|$ and $Q(a_N b_N)=1$, then the local boundedness of the system states is not guaranteed. In order to avoid this to happen, if $\wp \cap B_0 = \text{Ø}$, or if $\exists k' \in Z$ such that $|x'(k)| < |b_N|$ and $Q(a_N b_N)=1$, then we define

$$\mu_{\text{stable}}\left(\mathbf{x}^c(k_0+1)\right) \equiv \begin{cases} \delta_{\text{stable}} & \mathbf{x}^c(k_0+1)\in B_0 \\ 0 & \mathbf{x}^c(k_0+1)\in \Re^N \setminus B_0 \end{cases}, \tag{4.29}$$

where $1 \ge \delta_{\text{stable}} > 0$ and δ_{stable} is very close to zero. The reasons for why a small value of δ_{stable} can avoid the unboundedness problem of the system states will be discussed in Section 4.3.1. Since the fuzzy membership value of the system states outside B_0 is exactly equal to zero, this fuzzy membership function will force the new system states $\mathbf{x}^c(k_0+1)$ to be within B_0.

As discussed in Section 1.3.5, the occurrence of limit cycle behavior should be avoided. Since $\bigcup_{\forall P>0} \Psi_P$ is the set of system states that exhibits limit cycle behavior, we do not want to move the new system states $\mathbf{x}^c(k_0+1)$ into $\bigcup_{\forall P>0} \Psi_P$. Moreover, we do not want to move $\mathbf{x}^c(k_0+1)$ into $\bigcup_{\forall k \le k_0} \{\mathbf{x}(k)\}$ too. This is because after a certain number of

iterations, the system states may go to the same positions in the state space and cause limit cycle behavior. Define

$$PER(k_0) \equiv \left(\bigcup_{\forall P>0} \Psi_P \right) \cup \left(\bigcup_{\forall k \le k_0} \{\mathbf{x}(k)\} \right). \tag{4.30}$$

If $PER(k_0) \cap B_0 = B_0$, then all the system states in B_0 will result in limit cycle behaviors and this situation should be avoided. On the other hand, if $PER(k_0) \cap B_0 = \emptyset$, then we cannot find a system state $\mathbf{x}^q \in B_0 \cap PER(k_0)$ such that $\left\| \mathbf{x}(k_0+1) - \mathbf{x}^q \right\|_2$ is minimum. Hence, if $PER(k_0) \cap B_0 = B_0$ or $PER(k_0) \cap B_0 = \emptyset$, we define the fuzzy membership function as

$$\mu_{\text{aperiodic}}\left(\mathbf{x}^c(k_0+1)\right) \equiv \begin{cases} \delta_{\text{aperiodic}} & \mathbf{x}^c(k_0+1) \in B_0 \\ 0 & \mathbf{x}^c(k_0+1) \in \Re^N \setminus B_0 \end{cases}, \tag{4.31}$$

where $1 \ge \delta_{\text{aperiodic}} > 0$ and $\delta_{\text{aperiodic}}$ is also very close to zero. Similarly, the reason why a small value of $\delta_{\text{aperiodic}}$ can avoid the occurrence of limit cycle behavior will be discussed in Section 4.3.1. Otherwise, we define the fuzzy membership function as

$$\mu_{\text{aperiodic}}\left(\mathbf{x}^c(k_0+1)\right) \equiv \begin{cases} 1 - \left(\prod_{i=1}^{N} f_i\left(\mathbf{x}^c(k_0+1), \mathbf{x}^q\right) \right)^{\frac{1}{N}} & \mathbf{x}^c(k_0+1) \in B_0 \\ 0 & \mathbf{x}^c(k_0+1) \in \Re^N \setminus B_0 \end{cases}. \tag{4.32}$$

Since f_i is a triangular fuzzy membership function and $\mathbf{x}^q \in B_0$, $\mu_{\text{aperiodic}}\left(\mathbf{x}^c(k_0+1)\right) = 0$ when $\mathbf{x}(k_0+1) \in B_0 \cap PER(k_0)$ because $\mathbf{x}^q = \mathbf{x}(k_0+1)$ when $\mathbf{x}(k_0+1) \in B_0 \cap PER(k_0)$, $\mu_{\text{aperiodic}}\left(\mathbf{x}^c(k_0+1)\right) = 0$ when $\mathbf{x}^c(k_0+1) \in \Re^N \setminus B_0$ and $0 \le \mu_{\text{aperiodic}}\left(\mathbf{x}^c(k_0+1)\right) \le 1$ $\forall \mathbf{x}^c(k_0+1) \in B_0$. Hence, $\mu_{\text{aperiodic}}\left(\mathbf{x}^c(k_0+1)\right)$ forces the new system states $\mathbf{x}^c(k_0+1)$ to be within B_0. Since $\mu_{\text{aperiodic}}\left(\mathbf{x}^c(k_0+1)\right)$ captures the knowledge on the separation between $\mathbf{x}^c(k_0+1)$ and $B_0 \cap PER(k_0)$, which also reflects the separation between $\mathbf{x}^c(k_0+1)$ and the set of system states within the desired bounded region that exhibits a limit cycle behavior, $\mu_{\text{aperiodic}}\left(\mathbf{x}^c(k_0+1)\right)$ can be used to avoid the occurrence of limit cycle behaviors.

Once the fuzzy membership functions are defined, we can define the fuzzy impulsive control law as follows:

If $\mathbf{Ax}(k_0)+\mathbf{B}(\bar{\mathbf{u}}-Q(\mathbf{x}(k_0)))\in \Re^N \setminus B_0$, then the fuzzy impulsive controller will reset the state variables of the loop filter to $\mathbf{x}^c(k_0+1)$ where $\mathbf{x}^c(k_0+1)$ is the system state such that the following function is maximized,

$$\mu_{\mathbf{x}^c(k_0+1)}(\mathbf{x}^c(k_0+1)) \equiv \max_{\mathbf{x}^c(k_0+1)\in\Re^N}\left(\mu_{\text{stable}}(\mathbf{x}^c(k_0+1))\mu_{\text{aperiodic}}(\mathbf{x}^c(k_0+1))\mu_{\text{distance}}(\mathbf{x}^c(k_0+1))\right)^{1/3}. \quad (4.33)$$

Otherwise, no control force is applied to the SDM.

Lemma 5

$\forall \bar{u}\in\Re$, $\forall \mathbf{x}(0)\in\Re^N$, $\forall a_i\in\Re$ for $i=0,1,\cdots,N$ and $\forall b_j\in\Re$ for $j=1,\cdots,N$, $\mathbf{x}^c(k)\in B_0$ for $k>0$.

Proof: (please see Appendix E)

Different values of $\bar{u}$, $\mathbf{x}(0)$, a_i for $i=0,1,\cdots,N$ and b_j for $j=1,\cdots,N$, will affect the existence of $\wp$ and $\bigcup_{\forall P>0}\Psi_P$. However, Lemma 5 is still applied even though $\wp=\emptyset$ or $\wp=B_0$, and $\bigcup_{\forall P>0}\Psi_P=\emptyset$ or $\bigcup_{\forall P>0}\Psi_P=B_0$. Hence, Lemma 5 guarantees that the controlled trajectory is bounded within B_0 no matter what the input step size, the initial condition and the filter parameters are. It is very important because we do not want the trajectory of the SDM to be unbounded when either the input step size is increased, or the initial condition or the loop filter coefficients of the SDMs is changed. Another advantage of this fuzzy impulsive control strategy is that we can alter the maximum bound of the state variables easily by setting the value of V_{cc} appropriately, which is independent of the input step size, the initial condition and the filter parameters.

Lemma 6

$\forall \bar{u}\in\Re$, $\forall \mathbf{x}(0)\in\Re^N$, $\forall a_i\in\Re$ for $i=0,1,\cdots,N$ and $\forall b_j\in\Re$ for $j=1,\cdots,N$, $\left\|\mathbf{x}^c(k+1)-\mathbf{x}^r\right\|_2 \le 2V_{\text{cc}}\sqrt{N}$ for $k>0$.

Proof: (please see Appendix F)

The importance of this lemma is that it guarantees the norm of the difference between $\mathbf{x}^r$ and $\mathbf{x}^c(k+1)$ being bounded by $2V_{\text{cc}}\sqrt{N}$, no matter what the input step size, the initial condition and the filter parameters are. Since we do not want the norm of the

difference between $\mathbf{x}^r$ and $\mathbf{x}^c(k+1)$ to be too large because the effect of audible clicks which can be too large, this lemma helps us to estimate the worst case for the upper bound of the audible click effect.

Lemma 7

If $\exists k_0 \in Z$ such that $PER(k) \bigcap B_0 \neq B_0$ for $k \geq k_0$, and $\mathbf{Ax}(k_0) + \mathbf{B}(\overline{\mathbf{u}} - Q(\mathbf{x}(k_0))) \in \Re^N \setminus B_0$, then $\exists M > 0$ such that $\mathbf{x}^c(k) = \mathbf{x}^c(k+M)$ for $k > k_0$.

Proof: (please see Appendix G)

The importance of this lemma is that it states the condition that limit cycle behaviors do not occur when the fuzzy impulsive control strategy is applied at once. We will show in contrast to Section 4.4 that clipping usually results in limit cycle behavior, while our approach can avoid the occurrence of limit cycle behavior.

4.3.1 Parameters in the fuzzy impulsive controller

There are only three parameters in the fuzzy impulsive control strategy. They are V_{cc}, $\delta_{\text{aperiodic}}$ and δ_{stable}. V_{cc} is the maximum allowable bound on each state variable and this value is determined from the real situations, such as the hardware constraints and the safety specifications, etc. For example, if the hardware operates normally in a safe condition only when the state variables are bounded by 20V, then V_{cc} will be set accordingly. For the parameters $\delta_{\text{aperiodic}}$ and δ_{stable}, the fuzzy impulsive controller works properly $\forall \delta_{\text{aperiodic}} \in (0,1]$ and $\forall \delta_{\text{stable}} \in (0,1]$. However, since $\delta_{\text{aperiodic}}$ represents the fuzzy membership value of how to avoid the occurrence of limit cycle at $\mathbf{x}^c(k_0+1)$ when $PER(k_0) \bigcap B_0 = B_0$ or $PER(k_0) \bigcap B_0 = \emptyset$, and all the system states in B_0 may cause the trajectory to exhibit limit cycle behavior if $PER(k_0) \bigcap B_0 = B_0$, we suggest the SDM control designers setting this value to a small number, such as 10^{-3}, because this refers to the case which will not happen. For δ_{stable}, since it represents the fuzzy membership value of the local boundedness of the system states of the SDM at $\mathbf{x}^c(k_0+1)$ if $\wp \bigcap B_0 = \emptyset$, or if $\exists k' \in Z$ such that $|x'(k)| < |b_N|$ and $Q(a_N b_N) = 1$, the system state of the SDM will be unbounded if the fuzzy impulsive control strategy is not applied, we recommend the SDM control designers setting this value to a small number, for example, 10^{-3}.

4.3.2 Complexity issue

Although more fuzzy rules and sophisticated fuzzy engine can improve the performance of the SDM, they will increase the complexity of the system and may cause some real time processing problems, particularly in audio applications. This is because the Nyquist sampling rate for audio signal is 44.1kHz. Since the input signals are typically oversampled at 64 or 128, the number of samples inputted to the SDM per second is 2.8224M or 5.6448M. Since several mega samples need to be processed per second, only three basic fuzzy rules are captured and only a simple fuzzy engine is used to restrict the complexity in the processing. According to the simulation results shown in Section 4.4, three basic rules and a simple fuzzy engine are sufficient for achieving the objectives.

4.3.3 Implementation of the fuzzy impulsive controller

As discussed in the beginning of Section 5.3, the fuzzy impulsive controller resets the state variables of the loop filter to the controlled state variables $\mathbf{x}^c(k_0+1)$ if $\mathbf{A}\mathbf{x}(k_0)+\mathbf{B}(\bar{\mathbf{u}}-Q(\mathbf{x}(k_0)))\in \Re^N \setminus B_0$, where $\mathbf{x}^c(k_0+1)$ is calculated based on equation (4.33). Numerical solvers, such as MATLAB or MATHCAD, can be employed for solving equation (4.33). To reset the state variables of the loop filter, many existing reset circuits can be employed.

4.4 Simulation results

To illustrate our results, a fifth order SDM with loop filter transfer function

$$\frac{20z^{-1}-74z^{-2}+103.0497z^{-3}-64.0015z^{-4}+14.9584z^{-5}}{1-5z^{-1}+10.0025z^{-2}-10.0075z^{-3}+5.0075z^{-4}-1.0025z^{-5}} \tag{4.34}$$

is illustrated. This fifth order SDM was employed in [4]. The SDM can be implemented via the Jordan form [4] or can be realized as the following state space equation

$$\tilde{\mathbf{x}}(k+1)=\tilde{\mathbf{A}}\tilde{\mathbf{x}}(k)+\tilde{\mathbf{B}}(u(k)-y(k)) \tag{4.35}$$

for $k\geq 0$, where

$$y(k)=Q(\tilde{\mathbf{C}}\tilde{\mathbf{x}}(k)), \tag{4.36}$$

$$\tilde{\mathbf{A}} \equiv \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & -0.0018 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & -0.000685 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}, \tilde{\mathbf{B}} \equiv \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \text{ and } \tilde{\mathbf{C}} \equiv \begin{bmatrix} 20 \\ 6 \\ 1 \\ 0.09375 \\ 0.00589 \end{bmatrix}^T . \quad (4.37)$$

Assume that the initial condition of this SDM is zero, that is, $\tilde{\mathbf{x}}(0) = [0, \quad 0, \quad 0, \quad 0, \quad 0]^T$. By using a simple transformation, this SDM can be realized by the direct form and the corresponding initial condition is $\mathbf{x}(0) = [0, \quad -5, \quad 28.5, \quad 32.25, \quad 35.9793]^T$ when $u = 0.75$. We can check that the trajectory of this SDM will be bounded for this initial condition ($\tilde{\mathbf{x}}(0) = [0, \quad 0, \quad 0, \quad 0, \quad 0]^T$) if the input step size is approximately between -0.71 and 0.75, and will diverge if the input step size is outside this range. The relationship between the maximum absolute value of the state variables (realized in the direct form) and the input step size is plotted in Figure 4.2. From the simulation result, we can see that even though the trajectory is bounded for this range of input step size, the maximum absolute value of the state variables is between 20.0523 and 59.4633, which may be too large for some practical applications [4]. Figure 4.2 also shows the plot of the maximum absolute value of the state variables (also realized in the direct form) for $k > 0$ versus the input step size when the fuzzy impulsive control strategy is applied at $V_{cc} = 20$. According to Lemma 5, the maximum absolute value of the state variables of the controlled SDM is bounded by V_{cc} for $k > 0$ and $\forall \bar{u} \in \Re$, even though $|\bar{u}| \geq V_{cc}$. Hence, we can guarantee that the state variables are bounded by 20.

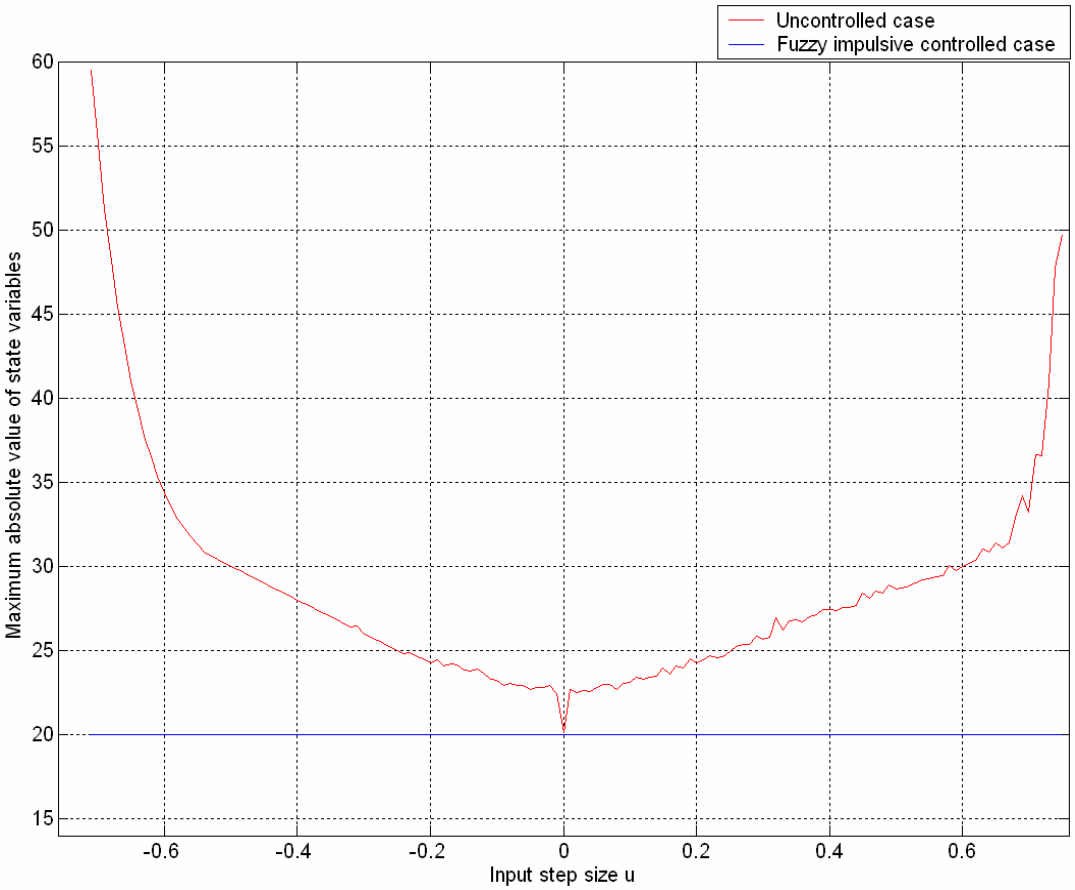


Figure 4.2 Plot of the maximum absolute value of the state variables (realized in direct form) against the input step size when $\tilde{\mathbf{x}}(0)=[0,\ \ 0,\ \ 0,\ \ 0,\ \ 0]^T$.

The global boundedness of the system states of this SDM is not valid. That means, $\exists\tilde{\mathbf{x}}(0)\in\Re^N$ such that the trajectory is unbounded. For example, when $\bar{u}=0.75$, Figure 4.3a and Figure 4.3b show the responses of $x_1(k)$ with $\tilde{\mathbf{x}}(0)=[0,\ \ 0,\ \ 0,\ \ 0,\ \ 0]^T$ and $\tilde{\mathbf{x}}(0)=[0.001,\ \ 0,\ \ 0,\ \ 0,\ \ 0]^T$, respectively. It can be seen from Figure 4.3a and Figure 4.3b that even though the SDM exhibits acceptable behavior when $\tilde{\mathbf{x}}(0)=[0,\ \ 0,\ \ 0,\ \ 0,\ \ 0]^T$ and the difference between these two initial conditions is very small, the system states of the SDM are unbounded when $\tilde{\mathbf{x}}(0)=[0.001,\ \ 0,\ \ 0,\ \ 0,\ \ 0]^T$ and the behaviors of the SDM for these two different initial conditions are very different. On the other hand, according to Lemma 5, the maximum absolute value of the state variables is always bounded by V_{cc} for $k>0$ and

$\forall \mathbf{x}(0) \in \Re^N$ if the fuzzy impulsive control strategy is applied. Figure 4.3c and Figure 4.3d show the corresponding state responses when the fuzzy impulsive control strategy is applied at $V_{cc} = 40$. From the simulation result, we see that the SDM exhibits acceptable behavior with the state variables bounded by V_{cc} for both of these two initial conditions.

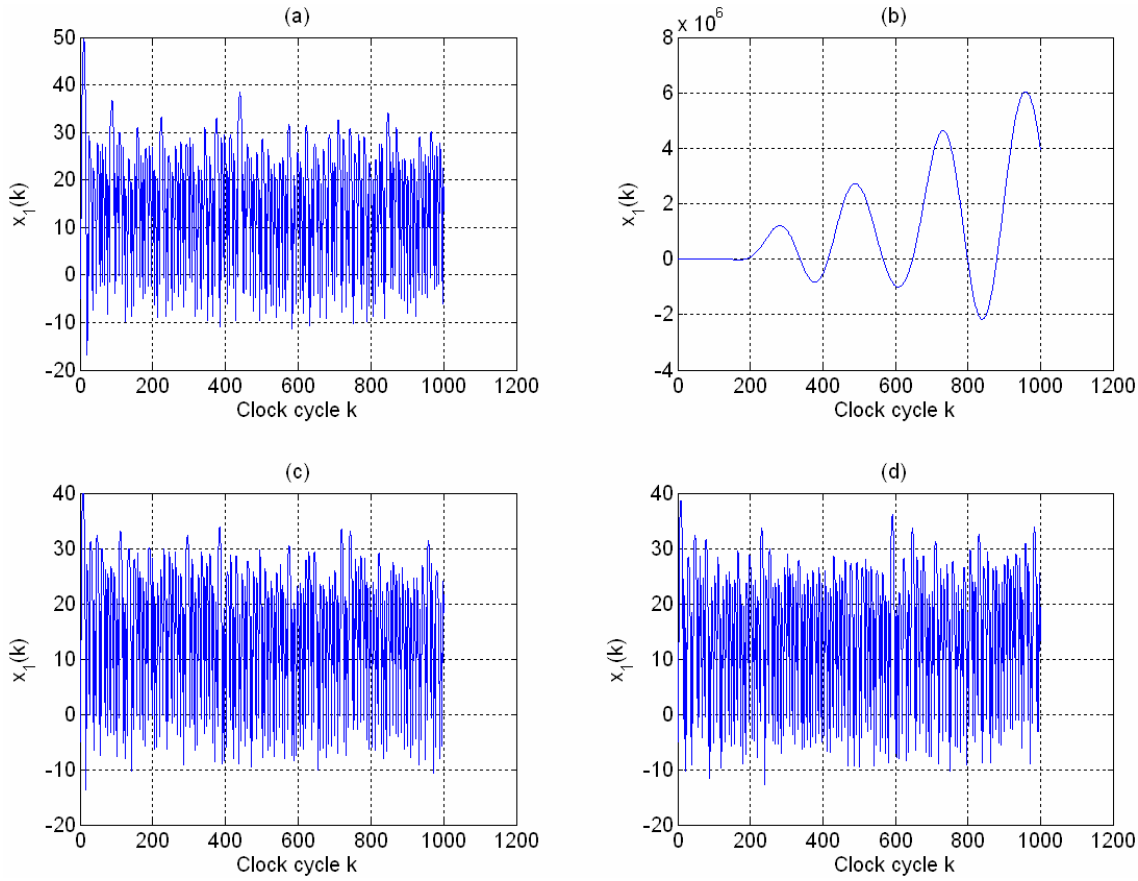


Figure 4.3 The response of $x_1(k)$ when $\bar{u} = 0.75$ and (a) initial condition $\tilde{\mathbf{x}}(0) = [0, \quad 0, \quad 0, \quad 0, \quad 0]^T$ when no control strategy is applied. (b) initial condition $\tilde{\mathbf{x}}(0) = [0.001, \quad 0, \quad 0, \quad 0, \quad 0]^T$ when no control strategy is applied. (c) initial condition $\tilde{\mathbf{x}}(0) = [0, \quad 0, \quad 0, \quad 0, \quad 0]^T$ when the fuzzy impulsive control strategy with $V_{cc} = 40$ is applied. (d) initial condition $\tilde{\mathbf{x}}(0) = [0.001, \quad 0, \quad 0, \quad 0, \quad 0]^T$ when the fuzzy impulsive control strategy with $V_{cc} = 40$ is applied.

Let us compare with other control strategies, such as the time delay feedback control strategy in which the controller is in the form $-K_c(1 - z^{-1})$. Denote λ_i for

$i = 1,2,\cdots,6$ as the poles of the effective loop filter. Since λ_i for $i = 1,2,\cdots,6$ depend on the value of K_c, it can be shown that $\max_{i=1,2,\cdots,6} |\lambda_i| > 1 \ \forall K_c \in \Re$ and the minimum value of $\max_{i=1,2,\cdots,6} |\lambda_i|$ occurs at $K_c = 0$. When $K_c = 0$, it reduces to the uncontrolled case. By selecting a value of K_c which is very close to zero, for example $K_c = 2 \times 10^{-5}$, and setting the initial condition and the input step size as the previous values, that is, $\mathbf{x}(0) = [0, \ -5, \ 28.5, \ 32.25, \ 35.9793, \ 39.5612]^T$ and $\bar{u} = 0.75$ (the initial condition is determined from the zero initial condition in Jordan form), it is found that the trajectory diverges and this is shown in Figure 4.4. This shows that the time delay feedback control strategy fails to stabilize the SDM.

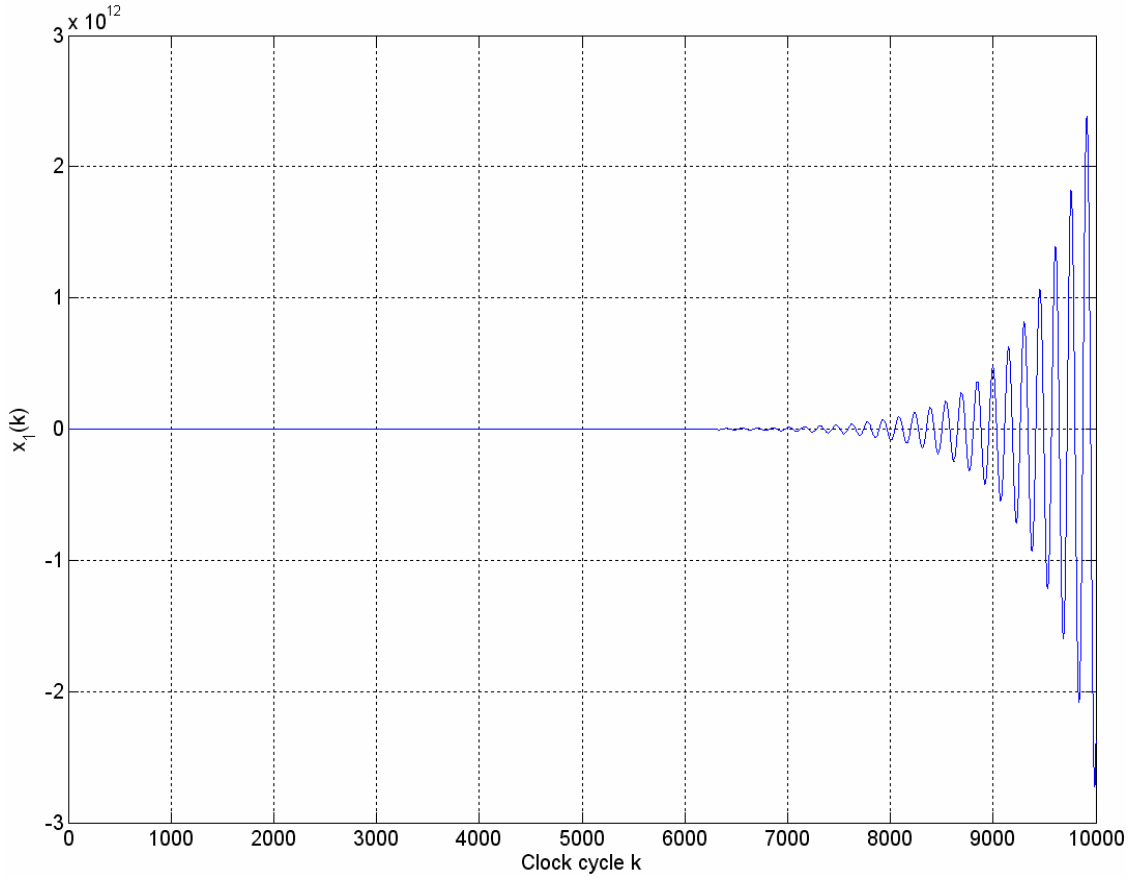


Figure 4.4 The response of $x_1(k)$ with input step size $\bar{u} = 0.75$ and initial condition $\mathbf{x}(0) = [0, \ -5, \ 28.5, \ 32.25, \ 35.9793, \ 39.5612]^T$ when the time delay feedback control strategy with $K_c = 2 \times 10^{-5}$ is applied.

To compare the fuzzy impulsive control strategy to the clipping control strategy, that is, set $x_i(k) = V_{cc} Q(x_i(k))$ whenever $|x_i(k)| \geq V_{cc}$ for $i = 1,2,\cdots,N$, it is found that limit cycle behaviors may occur if the clipping control strategy is applied. Figure 4.5 shows the magnitude response of $s(k)$ when $\bar{u} = 0.75$, $\tilde{\mathbf{x}}(0) = [0, \quad 0, \quad 0, \quad 0, \quad 0]^T$ and the clipped level is set at 40. It can be seen from Figure 5.5 that there is an impulse located at $\frac{\pi}{2}$ if the clipping control strategy is applied. This demonstrates that the SDM exhibits a limit cycle with period 2. On the other hand, the spectrum is relatively flat for the SDM when the fuzzy impulsive control strategy is applied with $V_{cc} = 40$. This demonstrates that the SDM exhibits normal behavior and the limit cycle behavior is avoided.

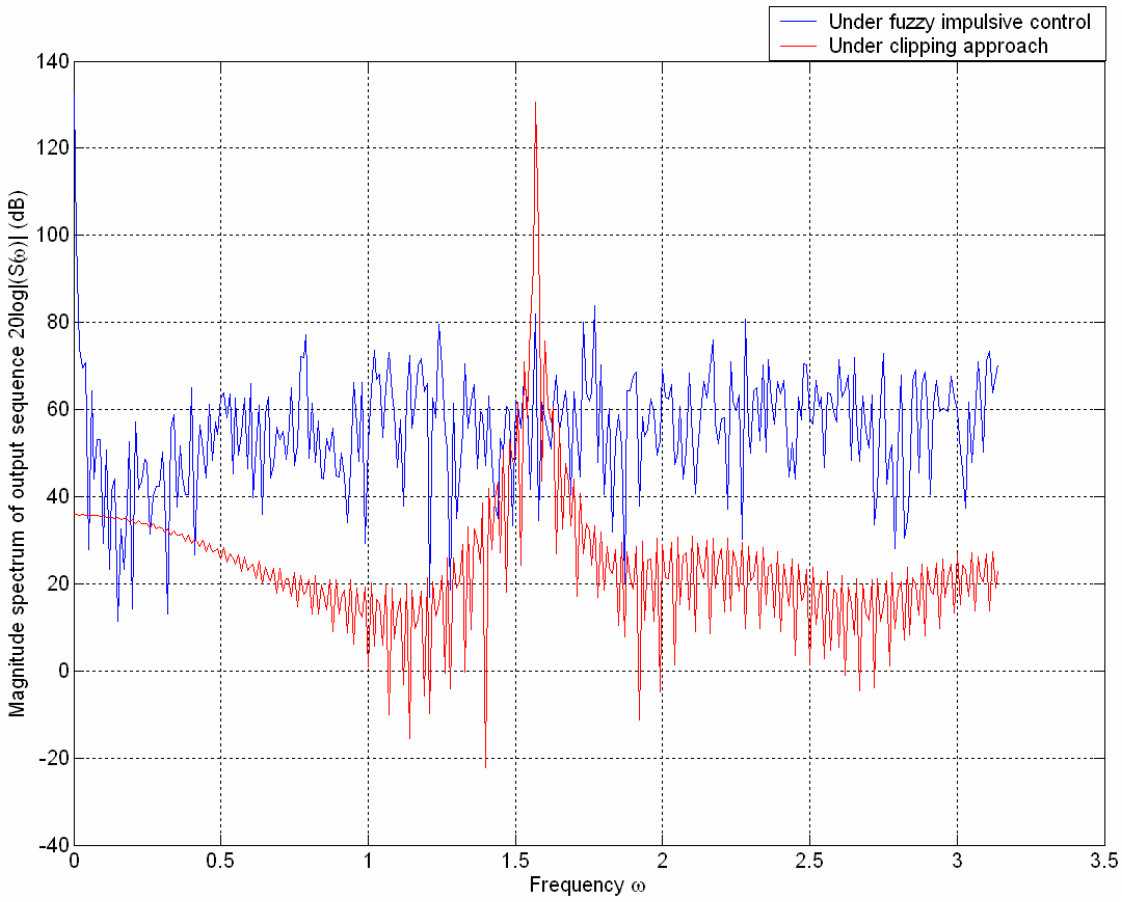


Figure 4.5 Comparison of the magnitude response of the output sequence when $\bar{u} = 0.75$ and initial condition $\tilde{\mathbf{x}}(0) = [0, \quad 0, \quad 0, \quad 0, \quad 0]^T$ for both the clipping and the fuzzy impulsive control strategies with the state variables bounded by 40.

Figure 4.6 shows the SNR of the SDM under clipping with the clipped level set to 28. SNR is calculated according to [74], where the frequency of the input sinusoidal signal is $\frac{2}{3}$ of its passband bandwidth. The OSR is 64, and the initial condition is given as $\tilde{\mathbf{x}}(0) = [0, \quad 0, \quad 0, \quad 0, \quad 0]^T$. It can be seen from Figure 4.6 that the SNR of both the SDMs with the clipping and fuzzy impulsive control strategies with the state variables bounded by 28 are the same when the input magnitude is less than 0.52. This is because both the maximum absolute value of the state variables (realized in the direct form) do not exceed 28. However, if the input magnitude is further increased, the SNR corresponding to the clipping control strategy will drop to less than 5dB as there is limit cycle behavior. On the other hand, the SDM exhibits stable behavior under the fuzzy impulsive control strategy. Hence, the SNR of the SDM under the fuzzy impulsive control strategy has an average of 30.5dB improvement compared to the clipping control strategy when the input magnitude is above 0.52.

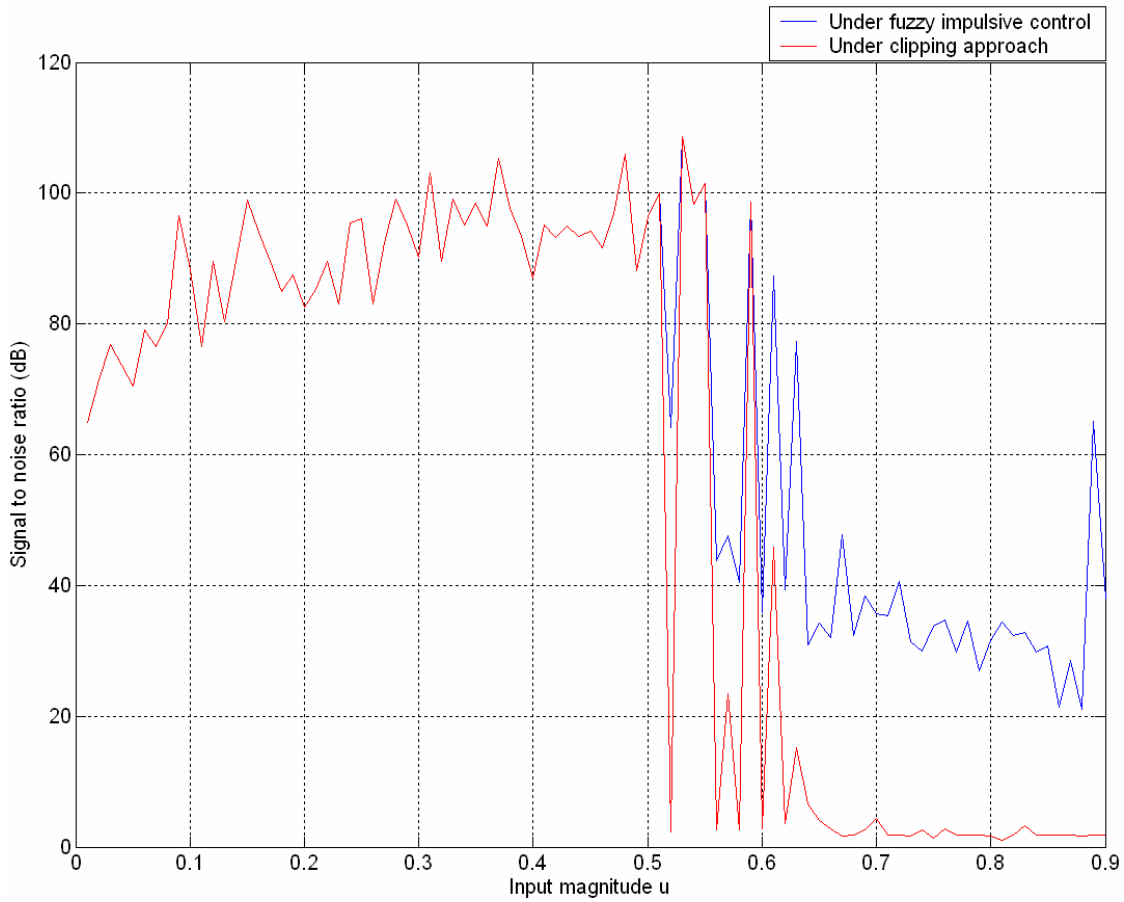


Figure 4.6 SNR of SDMs when input sinusoidal frequency is $\frac{2}{3}$ of the passband bandwidth, initial condition $\tilde{\mathbf{x}}(0) = [0, \quad 0, \quad 0, \quad 0, \quad 0]^T$ and the state variables are bounded by 28.

It can also be seen from Figure 4.7 that the probability of the control force being applied by the fuzzy impulsive control strategy is 0.011 for the input magnitude range greater than 0.52, as opposed to 0.7198 that for the clipping control strategy. Hence, the number of reset action on the state variables of the loop filter is much reduced when fuzzy impulsive control strategy is applied. This is because the fuzzy impulsive control strategy tends to reset the system states inside the invariant set if it exists and the system state will stay inside the invariant set without the need of applying the control force again. This demonstrates that the fuzzy impulsive control strategy is more efficient than the clipping control strategy.

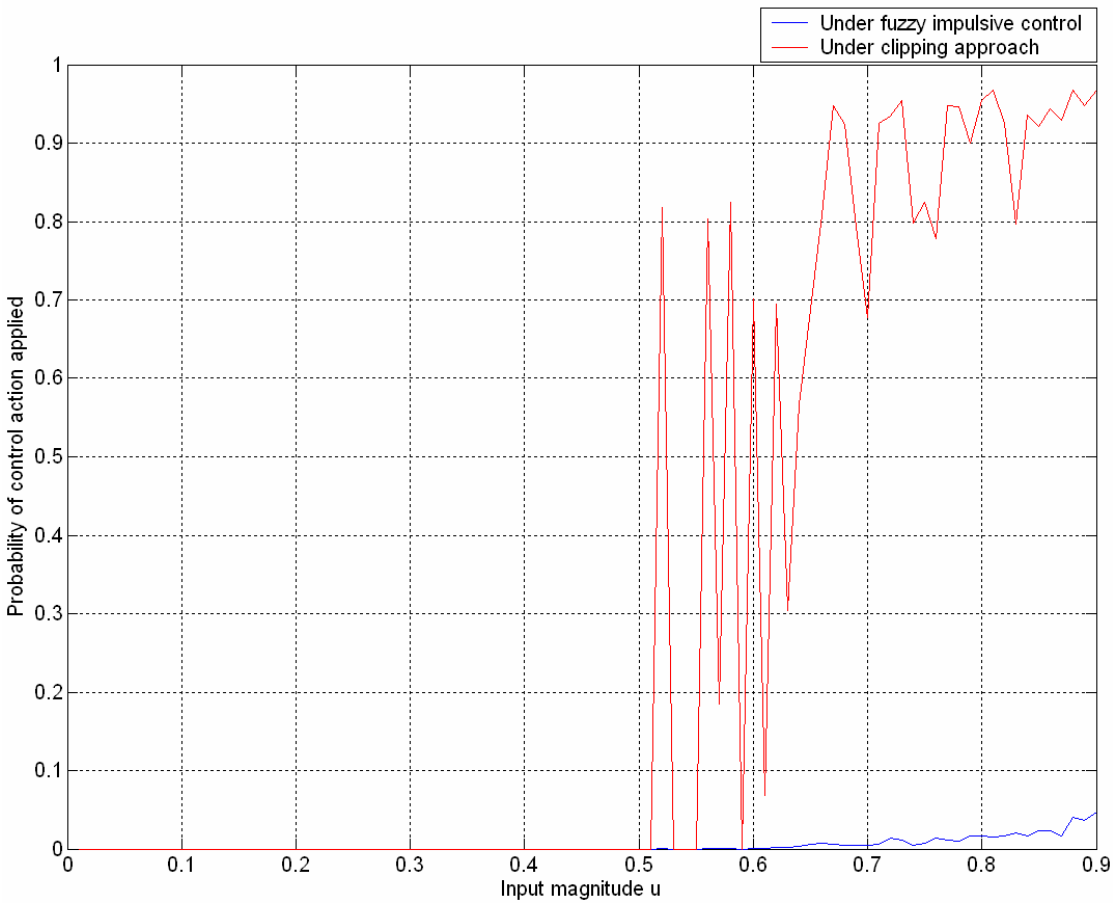


Figure 4.7 Probability of control force applied to the SDM when the input sinusoidal frequency is $\frac{2}{3}$ of the passband bandwidth, initial condition $\tilde{\mathbf{x}}(0)=[0,\quad 0,\quad 0,\quad 0,\quad 0]^T$ and the state variables are bounded by 28.

To verify the independence of the filter parameters on the fuzzy impulsive control strategy, consider another fifth order SDM with the following transfer function [4]

$$\frac{0.7919z^{-1}-2.8630z^{-2}+3.9094z^{-3}-2.3873z^{-4}+0.5498z^{-5}}{1-5z^{-1}+10.0023z^{-2}-10.0069z^{-3}+5.0069z^{-4}-1.0023z^{-5}}. \qquad (4.38)$$

This SDM has been used in the industry [4]. The trajectory of this SDM with $\bar{u}=0.59$ and $\tilde{\mathbf{x}}(0)=[0,\quad 0,\quad 0,\quad 0,\quad 0]^T$ is shown in Figure 4.8a, and it can be seen from Figure 4.8a that the trajectory diverges. On the other hand, when the fuzzy impulsive control strategy is applied with $V_{cc}=2$, according to Lemma 5, the maximum absolute value of

the state variables (realized in the direct form) is always bounded by V_{cc} for $k>0$, $\forall a_i \in \Re$ for $i=0,1,\cdots,N$ and $\forall b_j \in \Re$ for $j=1,\cdots,N$, as shown in Figure 4.8b.

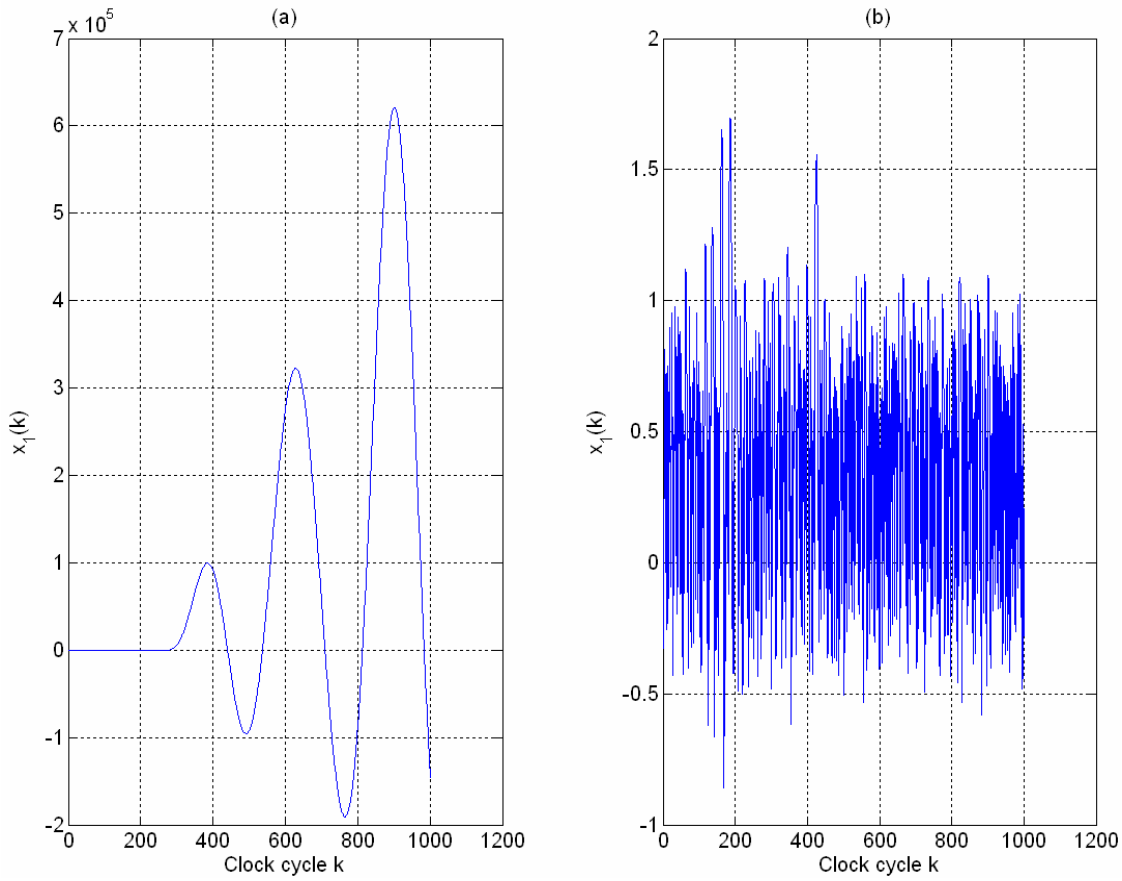


Figure 4.8 The response of $x_1(k)$ with initial condition $\tilde{\mathbf{x}}(0)=[0,\quad 0,\quad 0,\quad 0,\quad 0]^T$ and input step size $\bar{u}=0.59$ (a) when no control strategy is applied. (b) when the fuzzy impulsive control strategy with $V_{cc}=2$ is applied.

4.5 Conclusions

In this chapter, we have proposed the fuzzy impulsive control strategy for the stabilization of higher order interpolative SDMs in which the occurrence of limit cycle behaviors and the effect of audible clicks are minimized. Since the effective poles of the loop filter are not affected by the control strategy, the SNR performance of the SDM is maintained or improved after control. Since we perform learning and training for the whole state vector, the controlled trajectory is guaranteed to be bounded no matter what the input step size, the initial condition and the filter parameters are. This means that for

any deviation of the initial condition within the training set, the state will still be inside the invariant set. This gives a high robustness of the proposed controller.

Comparisons between the fuzzy impulsive control strategy and some existing control strategies show that the fuzzy impulsive control strategy is more effective in terms of producing much higher SNR and efficient in terms of requiring less number of control forces applied to the system.

CHAPTER V. DESIGN OF INTERPOLATIVE SDMS VIA SEMI-INFINITE PROGRAMMING

Since the phase information of a loop filter is sometimes not important and difficult to be characterized for some applications, such as audio applications [4], an IIR filter is chosen in preference to an FIR filter as the loop filter. Moreover, if we employ an FIR filter, then the order of the filter required will be very high and the computational problem will become significant. On the other side, by employing an IIR filter, we can be benefited from the lower passband and stopband ripples of the filter as well. However, since rational IIR filters consist of both the numerator and denominator coefficients, there are some challenges for designing rational IIR filters.

5.1 Issues for designing IIR filter and loop filters in SDMs

The major issue of designing the loop filter of an SDM is to achieve high SNR with the guarantee of the boundedness of state variables. Since the SDM consists of a quantizer, which is a nonlinear component, there is no simple relationship among the SNR, maximum bound of the input signal and the filter parameters, particularly when the filter order is high. Hence, it is typical to achieve high SNR by achieving good responses of both the STF and the NTF, and to achieve the boundedness of the state variables by keeping the stability conditions of the STF and the NTF. The objective of our formulated problem is to formulate an SDM design problem as two optimization problems based on the characteristics of the STF and the NTF, the stopband characteristics of loop filters and the stability conditions of the STF and the NTF. In order to achieve the boundedness of the system states, equation (10) in [73] directly implies that an IIR filter is stable if the real part of the denominator function is positive, that is, the sum of the numerator and denominator polynomials of the loop filter transfer function has to be on the right hand side of the complex plane for all frequencies.

5.2 SIP

Note that the noise shaping characteristics, as well as the frequency selectivity of the filters, are defined in the frequency domain, so all the constraints are continuous. Hence, the optimization problem is actually a quadratic SIP problem.

Since the solution is required to satisfy the constraints for all frequencies, simple methods for solving finite number of discrete constraint problems do not apply.

The most common methods for solving SIP problems are discretization methods, local reduction methods, dual exchange methods, nondifferentiable optimization approaches and interior point methods. For the discretization methods, it is not guaranteed that the continuous constraints are satisfied among the discretized points [73]. Although the difference between the exact upper bound of discretized constraints and that of the corresponding continuous constraints vanishes as the number of grid points increases, the computational complexity increases. For the local reduction methods, they require a good initial guess of a solution sufficiently close to the optimal solution in order to ensure its local convergence. For the dual exchange methods, they may have numerical instabilities, which means, this method is not robust to numerical quantization. For the nondifferentiable optimization approaches, they are not efficient to solve smooth problems. For the interior point methods, they are not applicable if the number of constraints tends to infinity.

5.3 Dual parameterization method

The dual parameterization method [57] is to parameterize the measure in the dual problem so that it transforms a SIP problem into equivalent finite dimensional nonlinear programming problem via sequence of regular convex programs. That is, from primal problem P to dual problem D:

Problem P

$$\min_{\mathbf{x}} J(\mathbf{x})$$

$$\text{subject to} \quad \mathbf{A}(\omega)\mathbf{x} \le \mathbf{C}(\omega) \qquad \text{where } \mathbf{x} \in \Re^N,\ \forall \omega \in \Omega.$$

Problem D

$$\min_{(\mathbf{x},\varpi)} J'(\mathbf{x},\varpi)$$

subject to $\mathbf{A}(\varpi_i)\mathbf{x} \le \mathbf{C}(\varpi_i)$ where $\varpi_i \in \Re$ and for $i = 1,2,\cdots,M$,

where M is the number of discrete frequencies, ϖ_i for $i = 1,2,\cdots,M$ are the discrete frequencies, and $\varpi = [\varpi_1, \cdots \varpi_M]^T$ is the frequency vector that is to be optimized. Problem P and problem D are equivalent in the sense that once the stationary points in problem P are determined, it is guaranteed that all ω in problem P would satisfy the constraints in problem P, but these stationary points are unknown, so we optimally determine these turning points and denote them as ω_i in problem D.

For the implementation of the dual parameterization method, first of all, we initialize a sequence of index set. Next we compute a local optimal solution by solving a finite dimensional nonlinear programming problem. Finally, we compute the global optimal solution via a local search for the finite dual problem.

With dual parameterization method, global optimal solution that satisfies the corresponding continuous constraint is guaranteed if the solution exists. The advantages of solving the SIP problem via the dual parameterization method is that the stability of the NTF and the STF are guaranteed. Moreover, it can be applied to design real, rational, causal loop filters without imposing specific filter structures such as Laguerre filter and Butterworth filter structures, and the avoidance of the iterative design [73] between the numerator and the denominator coefficients, which cannot be guaranteed for the convergence. Our simulation results show that this proposed design approach yields a significant improvement in the SNR compared to the existing design approaches.

5.4 Problem formulation

The design problem is formulated into two different optimization problems. The first optimization problem is to determine the denominator coefficients via minimizing the passband energy of the denominator of the loop filter transfer function (excluding the DC poles), subject to the continuous constraint on the maximum modulus square of the denominator of the loop filter transfer function. The second optimization problem is to determine the numerator coefficients via minimizing the stopband energy of the numerator of the loop filter transfer function, subject to the continuous constraint on the stability conditions of the NTF and the STF.

In this section, we consider the interpolative SDM shown in Figure 1.4. We only consider a lowpass SDM with at least one DC pole in the transfer function of the loop filter. The frequency response of the loop filter is assumed to be as follows:

$$H(\omega)=\frac{e^{-j\omega}\sum_{m=0}^{M}b_m e^{-jm\omega}}{\left(1-e^{-j\omega}\right)^r\left(1+\sum_{n=1}^{N}a_n e^{-jn\omega}\right)}, \tag{5.1}$$

where M and N are the numbers of roots of the polynomials of $e^{-j\omega}$ in the numerator and denominator of the transfer function of the loop filter (excluding the DC poles and pure delay elements), respectively, r is the number of DC poles, a_n, b_m for $n=1,2,\cdots,N$ and $m=0,1,\cdots,M$ are the filter coefficients. In the following consideration, $a_n, b_m \in \Re$, $r \geq 1$ and $N+r \geq M+1$. The design problem is equivalent to finding appropriate sets of filter coefficients a_n and b_m. Note that our design method can still be applied to the cases when the IIR filter is not causal or there is no DC pole in the transfer function.

By grouping the filter coefficients in the numerator and denominator as

$$\mathbf{x}_b \equiv \left[b_0, \quad \cdots, \quad b_M\right]^T \text{ and } \mathbf{x}_a \equiv \left[a_1, \quad \cdots, \quad a_N\right]^T, \tag{5.2}$$

respectively, and defining

$$\boldsymbol{\eta}_N(\omega) \equiv \left[1, \quad e^{-j\omega}, \quad \cdots, \quad e^{-jM\omega}\right]^T \tag{5.3}$$

and

$$\boldsymbol{\eta}_D(\omega) \equiv \left[e^{-j\omega}, \quad e^{-j2\omega}, \quad \cdots, \quad e^{-jN\omega}\right]^T, \tag{5.4}$$

then

$$H(\omega)=\frac{e^{-j\omega}\left(\boldsymbol{\eta}_N(\omega)\right)^T\mathbf{x}_b}{\left(1-e^{-j\omega}\right)^r\left(1+\left(\boldsymbol{\eta}_D(\omega)\right)^T\mathbf{x}_a\right)}. \tag{5.5}$$

The STF and NTF of the SDM can be expressed as

$$\mathrm{STF}(\omega)=\frac{H(\omega)}{1+H(\omega)} \text{ and } \mathrm{NTF}(\omega)=\frac{1}{1+H(\omega)}, \tag{5.6}$$

respectively. Denote the passband of the loop filter as B_P, which is the band of interest. For SDMs having a good SNR, the STF should be approximately equal to 1 and the NTF should be approximately equal to 0 for all frequencies in the passband of the loop filter.

This holds if $\left|1+\left(\boldsymbol{\eta}_D(\omega)\right)^T\mathbf{x}_a\right|\to 0 \ \ \forall\omega\in B_P$. Hence, we can define the cost function as follows:

$$\begin{aligned} J_a(\mathbf{x}_a) &\equiv \int_{B_P}\left|1+\left(\boldsymbol{\eta}_D(\omega)\right)^T\mathbf{x}_a\right|^2 d\omega \\ &= \int_{B_P}\left(1+\mathbf{x}_a{}^T\left(\boldsymbol{\eta}_D(\omega)\right)^*\right)\left(1+\left(\boldsymbol{\eta}_D(\omega)\right)^T\mathbf{x}_a\right)d\omega \\ &= \int_{B_P}\left(1+\left(\boldsymbol{\eta}_D(\omega)\right)^T\mathbf{x}_a+\left(\boldsymbol{\eta}_D(\omega)\right)^+\mathbf{x}_a+\mathbf{x}_a{}^T\left(\boldsymbol{\eta}_D(\omega)\right)^*\boldsymbol{\eta}_D(\omega)^T\mathbf{x}_a\right)d\omega \\ &= \int_{B_P}d\omega+2\left(\int_{B_P}\mathrm{Re}\left(\boldsymbol{\eta}_D(\omega)\right)^T d\omega\right)\mathbf{x}_a+\mathbf{x}_a{}^T\left(\int_{B_P}\left(\boldsymbol{\eta}_D(\omega)\right)^*\boldsymbol{\eta}_D(\omega)^T d\omega\right)\mathbf{x}_a \end{aligned} \quad , \tag{5.7}$$

where the superscript $^+$ denotes the transpose conjugate operator. Since

$$\begin{aligned} &\mathbf{x}_a{}^T\left(\boldsymbol{\eta}_D(\omega)\right)^*\boldsymbol{\eta}_D(\omega)^T\mathbf{x}_a \\ &=\left(\sum_{n=1}^N a_n e^{jn\omega}\right)\left(\sum_{n=1}^N a_n e^{-jn\omega}\right) \\ &=\left(\sum_{n=1}^N a_n\cos n\omega+\sum_{n=1}^N ja_n\sin n\omega\right)\left(\sum_{n=1}^N a_n\cos n\omega-\sum_{n=1}^N ja_n\sin n\omega\right) \\ &=\left(\sum_{n=1}^N a_n\cos n\omega\right)\left(\sum_{n=1}^N a_n\cos n\omega\right)-\left(\sum_{n=1}^N a_n\cos n\omega\right)\left(\sum_{n=1}^N ja_n\sin n\omega\right) \\ &\quad+\left(\sum_{n=1}^N ja_n\sin n\omega\right)\left(\sum_{n=1}^N a_n\cos n\omega\right)-\left(\sum_{n=1}^N ja_n\sin n\omega\right)\left(\sum_{n=1}^N ja_n\sin n\omega\right) \\ &=\left(\sum_{n=1}^N a_n\cos n\omega\right)\left(\sum_{n=1}^N a_n\cos n\omega\right)+\left(\sum_{n=1}^N a_n\sin n\omega\right)\left(\sum_{n=1}^N a_n\sin n\omega\right) \\ &=\mathbf{x}_a{}^T\,\mathrm{Re}\left(\boldsymbol{\eta}_D(\omega)\right)\mathrm{Re}\left(\boldsymbol{\eta}_D(\omega)\right)^T\mathbf{x}_a+\mathbf{x}_a{}^T\,\mathrm{Im}\left(\boldsymbol{\eta}_D(\omega)\right)\mathrm{Im}\left(\boldsymbol{\eta}_D(\omega)\right)^T\mathbf{x}_a \\ &=\mathbf{x}_a{}^T\left(\mathrm{Re}\left(\boldsymbol{\eta}_D(\omega)\right)\mathrm{Re}\left(\boldsymbol{\eta}_D(\omega)\right)^T+\mathrm{Im}\left(\boldsymbol{\eta}_D(\omega)\right)\mathrm{Im}\left(\boldsymbol{\eta}_D(\omega)\right)^T\right)\mathbf{x}_a \end{aligned} \quad , \tag{5.8}$$

the expression in (5.7) can be written in the following form

$$J_a(\mathbf{x}_a)=\frac{1}{2}\mathbf{x}_a{}^T\mathbf{Q}_a\mathbf{x}_a+\mathbf{b}_a{}^T\mathbf{x}_a+p_a\,, \tag{5.9}$$

which is a quadratic cost function, where

$$\mathbf{Q}_a\equiv 2\int_{B_P}\left(\mathrm{Re}\left(\boldsymbol{\eta}_D(\omega)\right)\mathrm{Re}\left(\boldsymbol{\eta}_D(\omega)\right)^T+\mathrm{Im}\left(\boldsymbol{\eta}_D(\omega)\right)\mathrm{Im}\left(\boldsymbol{\eta}_D(\omega)\right)^T\right)d\omega\,, \tag{5.10}$$

$$\mathbf{b}_a\equiv 2\int_{B_P}\mathrm{Re}\left(\boldsymbol{\eta}_D(\omega)\right)d\omega, \tag{5.11}$$

$$p_a \equiv \int_{B_P} d\omega, \tag{5.12}$$

and $\mathbf{Q}_a$ is a positive definite matrix.

Although the cost function minimizes the energy of the function $\left|1+\left(\boldsymbol{\eta}_D(\omega)\right)^T \mathbf{x}_a\right|$ over the passband of the loop filter, which reflects the error energy of the NTF over the passband, there may be a serious overshoot. If this case happens, then the SNR of the SDM will be degraded. To avoid this, a further constraint should be imposed, which bounds the function $\left|1+\left(\boldsymbol{\eta}_D(\omega)\right)^T \mathbf{x}_a\right|$ in the passband. The constraint is given by

$$\left|1+\left(\boldsymbol{\eta}_D(\omega)\right)^T \mathbf{x}_a\right|^2 \le \delta \quad \forall \omega \in B_P, \tag{5.13}$$

where δ denotes the bound. (5.13) can further be represented as

$$\frac{1}{2}\mathbf{x}_a{}^T \mathbf{A}_a(\omega)\mathbf{x}_a + \left(\mathbf{c}_a(\omega)\right)^T \mathbf{x}_a + q_a \le 0 \quad \forall \omega \in B_P, \tag{5.14}$$

where

$$\mathbf{A}_a(\omega) \equiv 2\left(\mathrm{Re}\left(\boldsymbol{\eta}_D(\omega)\right)\mathrm{Re}\left(\boldsymbol{\eta}_D(\omega)\right)^T + \mathrm{Im}\left(\boldsymbol{\eta}_D(\omega)\right)\mathrm{Im}\left(\boldsymbol{\eta}_D(\omega)\right)^T\right) \quad \forall \omega \in B_P, \tag{5.15}$$

$$\mathbf{c}_a(\omega) \equiv 2\,\mathrm{Re}\left(\boldsymbol{\eta}_D(\omega)\right) \quad \forall \omega \in B_P, \tag{5.16}$$

and

$$q_a \equiv 1-\delta \quad \forall \omega \in B_P. \tag{5.17}$$

Since $\mathbf{A}_a(\omega)$ is a positive definite matrix $\forall \omega \in B_P$ and the constraint is continuous, the design of the denominator coefficients can be formulated as the following SIP problem:

Problem (P_1)

$$\min_{\mathbf{x}_a} \qquad J_a\left(\mathbf{x}_a\right) = \frac{1}{2}\mathbf{x}_a{}^T \mathbf{Q}_a \mathbf{x}_a + \mathbf{b}_a{}^T \mathbf{x}_a + p_a, \tag{5.18a}$$

$$\text{subject to} \qquad \frac{1}{2}\mathbf{x}_a{}^T \mathbf{A}_a(\omega)\mathbf{x}_a + \left(\mathbf{c}_a(\omega)\right)^T \mathbf{x}_a + q_a \le 0 \quad \forall \omega \in B_P. \tag{5.18b}$$

Since the constraint function is convex in $\mathbf{x}_a$ and continuously differentiable with respect to both $\mathbf{x}_a$ and ω, the SIP problem can be solved by the dual parameterization method [57], which guarantees a global optimal solution that will satisfy the continuous quadratic constraint if the filter length is sufficiently long.

Although the characteristics of the NTF and STF are captured in the design, the stability of these two transfer functions and the frequency selectivity of the loop filter should also be considered. Our objective is to minimize the ripple energy of the loop filter in the stopband subject to the stability condition of both the STF and the NTF. Let the desired magnitude response of the loop filter be $\tilde{H}(\omega)$. In order to have good frequency characteristics for the loop filter, we want to achieve $|H(\omega)| \approx \tilde{H}(\omega)$, which implies that

$$\frac{\left|e^{-j\omega}\right|^2 \left|(\boldsymbol{\eta}_N(\omega))^T \mathbf{x}_b\right|^2}{\left|\left(1-e^{-j\omega}\right)^r\right|^2 \left|1+(\boldsymbol{\eta}_D(\omega))^T \mathbf{x}_a\right|^2} \approx \left|\tilde{H}(\omega)\right|^2$$

$$\left|(\boldsymbol{\eta}_N(\omega))^T \mathbf{x}_b\right|^2 \approx \left|\frac{\left(1-e^{-j\omega}\right)^r \tilde{H}(\omega)}{e^{-j\omega}}\right|^2 \left|1+(\boldsymbol{\eta}_D(\omega))^T \mathbf{x}_a\right|^2. \tag{5.19}$$

Since $\tilde{H}(\omega)$ is zero in the stopband, the cost function can be formulated as

$$J_b(\mathbf{x}_b) \equiv \int_{B_S} \left|(\boldsymbol{\eta}_N(\omega))^T \mathbf{x}_b\right|^2 d\omega, \tag{5.20}$$

where B_S denotes the stopband of the loop filter. According to equation (10) in [73], the stability condition of the NTF and STF is

$$\mathrm{Re}\left(e^{-j\omega}(\boldsymbol{\eta}_N(\omega))^T \mathbf{x}_b + \left(1-e^{-j\omega}\right)^r \left(1+(\boldsymbol{\eta}_D(\omega))^T \mathbf{x}_a\right)\right) \geq 0 \quad \forall \omega \in [-\pi, \pi], \tag{5.21}$$

which is equivalent to

$$\mathrm{Re}\left((\cos\omega - j\sin\omega)(\mathrm{Re}(\boldsymbol{\eta}_N(\omega)) - j\,\mathrm{Im}(\boldsymbol{\eta}_N(\omega)))^T \mathbf{x}_b + \left(1-e^{-j\omega}\right)^r \left(1+(\boldsymbol{\eta}_D(\omega))^T \mathbf{x}_a\right)\right) \geq 0 \quad \forall \omega \in [-\pi, \pi]$$

$$\mathrm{Re}(\cos\omega\,\mathrm{Re}(\boldsymbol{\eta}_N(\omega)) - \sin\omega\,\mathrm{Im}(\boldsymbol{\eta}_N(\omega)) - j\sin\omega\,\mathrm{Re}(\boldsymbol{\eta}_N(\omega)) - j\cos\omega\,\mathrm{Im}(\boldsymbol{\eta}_N(\omega)))\mathbf{x}_b + \mathrm{Re}\left(\left(1-e^{-j\omega}\right)^r \left(1+(\boldsymbol{\eta}_D(\omega))^T \mathbf{x}_a\right)\right) \geq 0 \quad \forall \omega \in [-\pi, \pi]$$

$$(\cos\omega\,\mathrm{Re}(\boldsymbol{\eta}_N(\omega)) - \sin\omega\,\mathrm{Im}(\boldsymbol{\eta}_N(\omega)))\mathbf{x}_b + \mathrm{Re}\left(\left(1-e^{-j\omega}\right)^r \left(1+(\boldsymbol{\eta}_D(\omega))^T \mathbf{x}_a\right)\right) \geq 0 \quad \forall \omega \in [-\pi, \pi]$$

$$(\boldsymbol{\eta}'_N(\omega))^T \mathbf{x}_b + \mathrm{Re}\left(\left(1-e^{-j\omega}\right)^r \left(1+(\boldsymbol{\eta}_D(\omega))^T \mathbf{x}_a\right)\right) \geq 0 \quad \forall \omega \in [-\pi, \pi], \tag{5.22}$$

where

$$\boldsymbol{\eta}'_N(\omega) \equiv [\cos\omega, \quad \cos 2\omega, \quad \cdots, \quad \cos(M+1)\omega]^T. \tag{5.23}$$

Hence, the optimization problem can be represented as the following SIP problem:

Problem (P_2)

$$\min_{\mathbf{x}_b} \quad J_b(\mathbf{x}_b) = \frac{1}{2}\mathbf{x}_b{}^T \mathbf{Q}_b \mathbf{x}_b, \tag{5.24a}$$

$$\text{subject to} \quad \mathbf{A}_b(\omega)\mathbf{x}_b + \mathbf{c}_b(\omega) \le 0 \quad \forall \omega \in [-\pi, \pi], \tag{5.24b}$$

where

$$\mathbf{Q}_b \equiv 2\int_{B_S} \left(\mathrm{Re}(\boldsymbol{\eta}_N(\omega))\mathrm{Re}(\boldsymbol{\eta}_N(\omega))^T + \mathrm{Im}(\boldsymbol{\eta}_N(\omega))\mathrm{Im}(\boldsymbol{\eta}_N(\omega))^T\right) d\omega, \tag{5.25}$$

$$\mathbf{A}_b(\omega) \equiv -(\boldsymbol{\eta}'_N(\omega))^T \quad \forall \omega \in [-\pi, \pi], \tag{5.26}$$

and

$$\mathbf{c}_b(\omega) \equiv -\mathrm{Re}\left(\left(1 - e^{-j\omega}\right)^r \left(1 + (\boldsymbol{\eta}_D(\omega))^T \mathbf{x}_a\right)\right) \quad \forall \omega \in [-\pi, \pi]. \tag{5.27}$$

Problem P_1 does not depend on the numerator coefficients, so the global optimal solution of problem P_1 can be obtained via the dual parameterization method [57]. Since the denominator coefficients are obtained from solving problem P_1, the global optimal solution of problem P_2 can then be obtained similarly. In this formulation, iterative design of the numerator and denominator coefficients is avoided. This is advantageous because convergence of the iterative design is usually not guaranteed.

5.5 Simulation results

To compare our design with the existing optimal designs, similar cost function and constraints should be used. However, few of them have the exactly same cost function and constraints. The most related existing design approach is the one based on the Butterworth filter structure or the Chebyshev filter structure because these two design methods employ SNR as the performance criterion.

Consider a fifth order SDM with a DC pole and a pure delay multiplied in the numerator of the loop filter transfer function and an OSR of 64, that is, $M = 5$ $N = 4$, $r = 1$, $B_P = \left[-\frac{\pi}{64}, \frac{\pi}{64}\right]$ and $B_S = [-\pi, \pi] \backslash \left[-\frac{\pi}{64}, \frac{\pi}{64}\right]$. We choose this configuration because this order of SDM and this value of OSR are typical in many audio applications [4]. It can be seen from Figure 5.1 that the maximum bound on $\left|1 + (\boldsymbol{\eta}_D(\omega))^T \mathbf{x}_a\right|$ for the design using the Chebyshev filter structure [74] is 1.9101×10^{-12}, while the design using the Butterworth structure is 1.5203×10^{-12}. Hence, we would expect that our result should

achieve the error bounded by 10^{-12} $\forall \omega \in B_P$. By selecting $\delta = 10^{-12}$, the optimal SDM design problem can now be formulated as SIP problems and these problems can be solved via the dual parameterization method [57]. According to the simulation, it is found that our design can achieve the error bounded by 9.6658×10^{-13}, as shown in Figure 5.1. It is worth noting that the design based on the Chebyshev filter structure or the Butterworth filter structure has a larger response value on the first lobe, while our design has a larger value on the second lobe. This implies that our design has a higher ability to shape the noise towards the high frequency band compared to the existing designs.

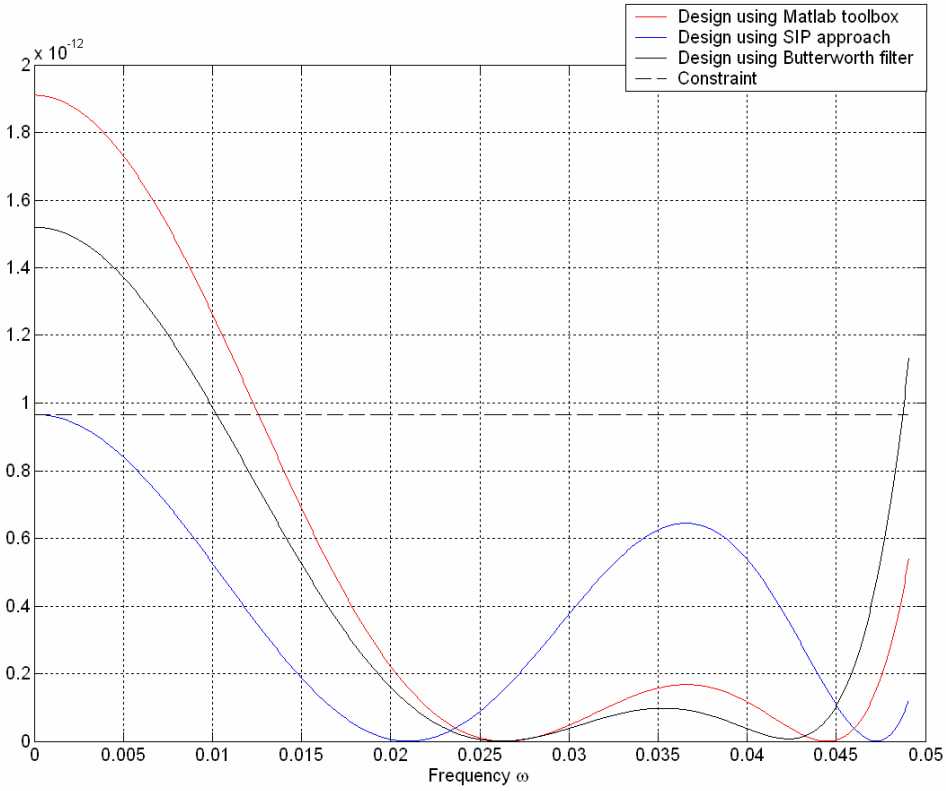


Figure 5.1 Comparison of the noise energies $\left|1+(\boldsymbol{\eta}_D(\omega))^T \mathbf{x}_a\right|^2$ of the passbands of the magnitude responses of different design approaches.

Figure 5.2 shows the SNRs of our design and the optimal designs using the Butterworth filter structure and the Chebyshev filter structure [74] based on sinusoidal inputs with input frequency equal to $\frac{2}{3}$ of the passband bandwidth. It can be seen from Figure 5.2 that our design can achieve an average of 5.0187dB improvement compared to that of [74] and 3.6639dB improvement compared to that of using Butterworth filter

structure when these SDMs operate normally. It is also worth noting that the system states of the design via the Chebyshev filter structure [74] diverges when the input sinusoidal magnitude is over 0.67, and that of the design via the Butterworth structure diverges at 0.61, while our design operates normally before 0.69. From the comparison, we can see that our design sustains a higher SNR as well as a higher input upper bound for bounded system states. It is found that the magnitude of the poles of the STF and the NTF of our design are 0.9928, 0.9928, 0.8556, 0.8556 and 0.6143, respectively, in which all are strictly inside the unit circle. Hence, the transfer functions are strictly stable.

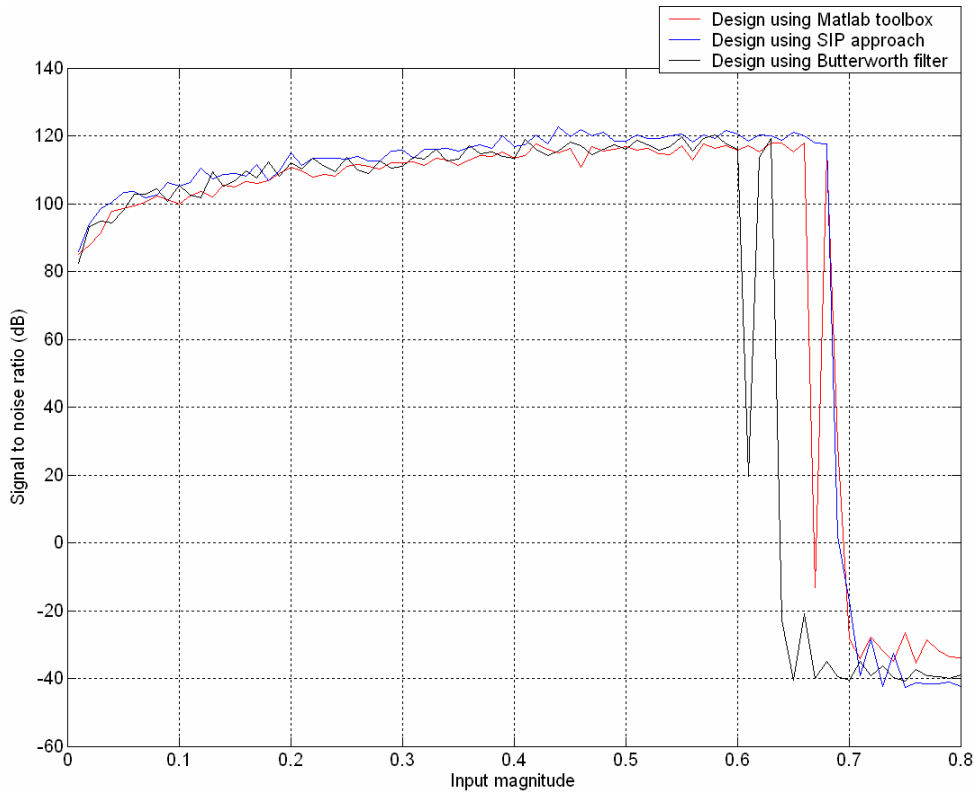


Figure 5.2 Comparison of the SNRs among different design approaches.

Figure 5.3 shows the NTFs of our proposed design, the design via the Chebyshev filter structure [74] and the design via the Butterworth filter structure with the OSR at 64 and Nyquist sampling rate at $44.1k$Hz [4]. According to the simulation results, we can see that our design produces at least 9.0146dB improvement on the passband of the loop filter in comparison to the design in [74] and 4.6748dB improvement in comparison to the design using Butterworth filter structure. These are significant improvements on the suppression of the noise on the frequency band we are interested.

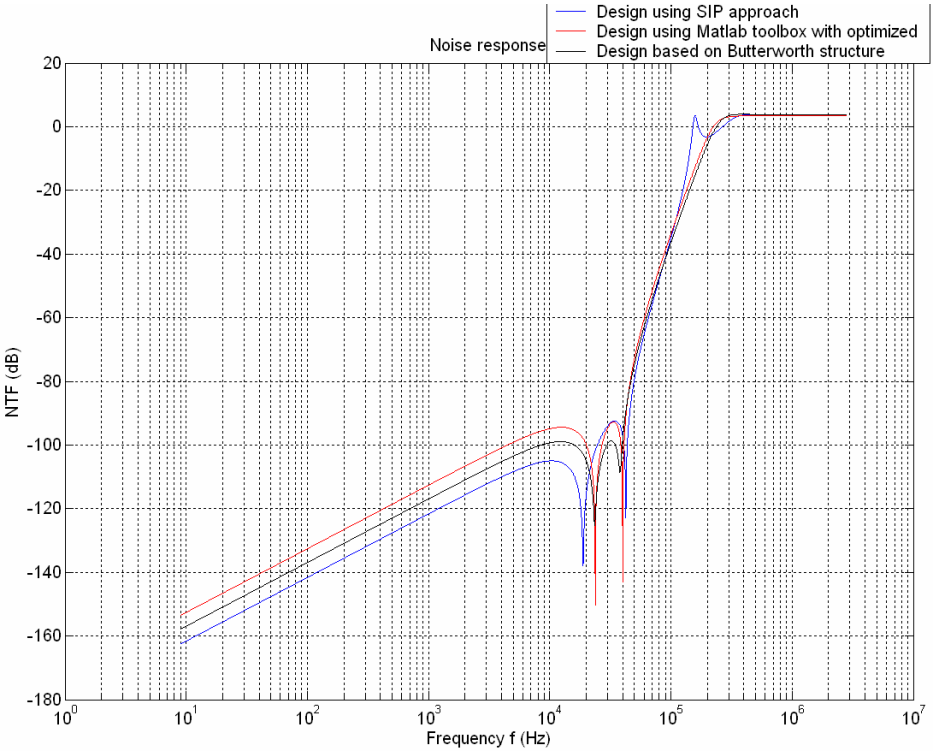


Figure 5.3 Comparison of the NTFs among different design approaches.

5.6 Conclusions

In this chapter, we have formulated SDM design problems as SIP problems and solved the problems via the dual parameterization method. The advantages of this formulation are the guarantee of the stability of NTF and STF given that the filter length is sufficiently long, applicability to design rational IIR filters without imposing specific filter structures such as Laguerre filter and Butterworth filter structures, and the avoidance of the nonconvergent iterative design of the numerator and the denominator coefficients. Our simulation results show that the proposed design yields a significant improvement in the SNR and achieves a higher stable input range compared to the existing designs.

CHAPTER VI. CONCLUSIONS AND FUTURE RESEARCH

In the thesis, the analysis, design and control of SDM is a very important process both to ensure robustness of the system to the environment and good stability and SNR performance. From engineering point of view, the analysis process determines the effect of the change of parameters such as the number of bits of the quantizer. The SDM usually employs a single bit quantizer that allows a fast operation for the system. We found that the existence of multibit quantizer in the SDM can be an important role because SDMs with single bit quantizer is not robust to noise after we treat the noise as an input to the system. One may consider the use of a multibit quantizer when the additive white guassian noise level is significantly high. The design process involves some subjective factors that the SDM is difficult to be optimized. The noise is a subjective factor to the performance. The control process must satisfy different objectives that possess an intelligent control to the system. From mathematical point of view, the SDM can be described by state space model which involves the mathematics of advanced linear algebra. The state space model allows us to consider the poles of the system that decides the stability of the system and the SNR performance as well. Overally, the research involves important issues from engineering and mathematical aspects.

The main objective of this thesis is to raise up the performances of SDM. To achieve this goal, nonlinear analysis (Chapter 2 and Chapter 3), control (Chapter 4) and design (Chapter 5) of the SDMs are the three common directions. In Chapter 2 and 3, we derived the state space equations for the SDM so that a better performance can be obtained in a chaotic but stable regime (Chapter 2 and 3) and the control (Chapter 4) and design (Chapter 5) objectives could be achieved. Chapter 2 showed that the SDM can be operated in chaotic regime, while Chapter 3 derived the conditions for the SDM to achieve chaotic but stable and high performance. When the conditions for a chaotic but stable operation of the SDM is not valid, we need to perform control (Chapter 4) that can force the SDM to operate in chaotic but stable regime. In Chapter 4, we perform the control for the SDM with existing filters. In Chapter 5, we design the filter for the SDM by numerical approach.

Some new results on multi-bit bandpass SDMs have been presented. It is found that elliptic fractal pattern of a bandpass SDM may be exhibited on the phase plane even though the saturation region of the quantizer is not activated and a high bit quantizer is used. Although we assume a constant input to the bandpass SDM, in real situation, the input signal is composed of noise that we cannot ignore. So when the number of bits of the quantizer increases, we will expect that the output of the quantizer is corrupted with noise. However, in this investigation, we show that a deterministic but nonlinear complex behavior is possible to occur so that the SDM operates normally. This allows us to avoid the effect of noise from the input if we can utilize this properly. This means that we will be able to avoid unwanted behavior to occur if we force the state of the SDM to be elliptic fractal state. In addition, we have found that a bit change in the quantizer can change the behavior of the system dramatically. For example, a bandpass SDM can change the behavior from limit cycle to elliptic fractal with only a single bit change of its quantizer. Therefore, intuitively, we are able to operate the bandpass SDM normally or even boost up the performance of the bandpass SDM by allowing a bit change of the quantizer.

Limit cycle behavior and unstable behavior directly affect the SDM performance. The conditions for avoiding limit cycle to occur and maintaining the overall system stability have been developed in Section 3.2 in Chapter 3. The optimal impulsive control problem, in which to minimize the change of the state before and after the impulsive control force is applied, subject to the avoidance of limit cycle behaviour and stability condition, has been developed in Section 3.2 in Chapter 3. From the understanding point of view of the problem formulation, further constraints can be imposed to the optimal impulsive control problem so that the controlled state can virtually fully satisfies the control objectives. For examples, clipping control cannot fully satisfy the control objectives when the controlled state is determined. However, by determining the forward dynamics and backward dynamics of the controlled system, in which these dynamics are defined in Section 3.2 in Chapter 3, then although at the moment of determining the controlled state the clipping control cannot fully satisfy the control objectives, there is space to explore for the existence and uniqueness of these dynamics. In this stage, the

forward and backward dynamics of clipping control will be investigated to justify their applications to sigma delta modulators.

We have accounted for the occurrence of near fractal and near chaotic patterns for a bandpass SDM with strictly stable system matrices. If the period of the limit cycles is larger than the difference of the phase portrait between the near fractal and the real fractal behaviors, or the difference between the near chaotic and the real chaotic behaviors are visually indistinguishable, then near fractal and near chaotic patterns will occur. Some interesting results have been found. First, for a periodic output sequence, the steady state period of the state space variables must be equal to the period of the symbolic sequence. This implies that all the periodic points cannot be in the same quadrant. If the system state converges to some fixed points on the phase portrait, then these fixed points will depend on the initial condition indirectly. One implication of these results is that we can generate signals with rich frequency spectra by using strictly stable system matrices in order to suppress unwanted tones generated by their quantizers with the guarantee of the bounded system states.

We have proposed a fuzzy impulsive control strategy for the stabilization of higher order interpolative SDMs in audio applications. The main advantage of this control strategy is that the effective poles of the loop filter are not affected, and so the SNR performance of the SDMs is maintained or improved after the control. The controlled trajectory is guaranteed to be bounded no matter what the input step size is and what the filter parameters and the initial conditions are. The bounded region can also be altered easily. Examples of higher order interpolative SDMs have been given to demonstrate the effective performance of the proposed control strategy. The main implication of fuzzy impulsive control is to achieve a bounded state trajectory and avoid the occurrence of limit cycles simultaneously, in which linear control strategies usually fail to stabilize the system for all initial conditions, input and system parameters, and simple nonlinear control strategies usually result in the occurrence of limit cycles. Impulsive control can stabilize the system by changing the state instead of changing the system. Once the state of the system goes to the invariant set, the system is guaranteed to be stable forever. Although the invariant set is well defined, the control rules for the

impulsive control force are fuzzy. Therefore, fuzzy impulsive control is proposed to determine where the state should be set so that the system performance can be sustained.

We have also formulated the SDM design problem as SIP problems and solve the problems via the dual parameterization method. The advantages of this formulation are the guarantee of the stability of the NTF and the STF if the solution exists, applicability to the design of rational IIR filters without imposing specific filter structures such as Laguerre filter and Butterworth filter structures, and the avoidance of the iterative design of numerator and the denominator coefficients, because the convergence of the iterative design is not guaranteed. Our simulation results show that the proposed design yields a significant improvement in the SNR compared to the existing designs. The main implication from the numerical simulation of the SIP design technique is that higher SNR can be achieved if continuous constraints are imposed on the stability of the NTF and the STF, as well as the frequency selectivity of the filter.

When we design the SDM, in order to achieve the boundedness of the system states, the real part of the sum of the numerator and denominator polynoimials of the loop filter transfer function has to be positive for all frequencies. We can consider the avoidance of the stability and the unwanted nonlinear behaviors of the SDM by including the conditions which was defined in Section 3.3 in Chapter 3. Then the optimization Problem (P_1) in equation (5.18) can be reformulated as follows.

Problem (P_1)

$$\min \quad J = J_a(\mathbf{x}_a) + f_i\left(\mathbf{x}^c(k_0+1), \mathbf{x}^r\right)$$

$$\text{where } J_a(\mathbf{x}_a) = \frac{1}{2}\mathbf{x}_a{}^T\mathbf{Q}_a\mathbf{x}_a + \mathbf{b}_a{}^T\mathbf{x}_a + p_a$$

and $f_i\left(\mathbf{x}^c(k_0+1), \mathbf{x}^r\right)$ which is defined in (5.27)

$$\text{subject to} \quad \frac{1}{2}\mathbf{x}_a{}^T\mathbf{A}_a(\omega)\mathbf{x}_a + (\mathbf{c}_a(\omega))^T\mathbf{x}_a + q_a \le 0 \quad \forall \omega \in B_P .$$

$$\text{and } \mathbf{x}^c(k_0+1) \le 1 - \left(\prod_{i=1}^{N} f_i\left(\mathbf{x}^c(k_0+1), \mathbf{x}^q\right)\right)^{\frac{1}{N}}$$

$$\text{where } \mathbf{x}^c(k_0+1) \in \Re^N \setminus B_0$$

We summarized our works as follows.

In the existing works on the occurrence of elliptic fractal patterns, it was shown in single bit bandpass SDMs only. In our work, we further showed that it does occur in multi-bit bandpass SDMs. Although a similar work was showed for the digital filter with two's complement arithmetic, it referred to the state when it changes from finite state machine to infinite state machine, while it referred to the state when it changes from lower bit quantizer to higher bit quantizer for the case of unsaturated bandpass SDMs. This is an interesting phenomena we have explored.

In the existing works, unstable poles are usually placed in SDMs to achieve fractal and chaotic behaviors to suppress unwanted tones from limit cycle behaviors. In our works, we have shown that near fractal or near chaotic signal can be generated by placing strictly stable poles to the SDMs. We can also guarantee the global boundedness of the system states.

In the existing works, the stable behaviors of the SDMs are usually controlled by strictly changing the systems, that is, changing the effective poles of the loop filters. It was shown that it is possible to control the SDMs without changing the systems by simply resetting the state each time when the states become unbounded. However, this led to the occurrence of unwanted tones. In our work, we proposed a more sophisticated approach. We only have to reset the states once when the states become unbounded to control the global boundedness of the states of the SDMs.

In the existing works, certain structures of the loop filters were proposed in the sub-optimal designs. In our works, we formulated the design problem as two optimization problems and we solved these SIP problems via dual parameterization approach. We showed that the performance of the SDMs can be greatly improved.

The shortcomings of my results are that it takes an infinite number of iterations to determine whether a point is in the invariant set defined in (4.24) or not. This is not practical. Hence, one possible future extension of the current work is to develop an efficient algorithm for characterising the invariant set.

One more possible future extension of the current work on design of sigma delta modulation is to design a bank of sigma delta modulators. By modulating the input signal to different frequency bands and designing a bank of filters so that the signal transfer

functions and the noise transfer functions are separated. As different frequency bands are exploited, higher SNR may be achieved.

Another possible future research is to generalize the stability of the symbolic dynamical systems. In our works, I only focus on the stability study of sigma delta modulators. Besides, we will look at the locations of the centre of the elliptical fractal regions.

APPENDIX A

For (i) implies (ii), since

$$\forall p, M \in Z^+ \text{ and } \forall k \geq 0,\ \mathbf{x}(k+pM) = \mathbf{A}^{pM}\mathbf{x}(k) + \sum_{n=0}^{pM-1} \mathbf{A}^{pM-1-n}\mathbf{B}(\mathbf{u} - \mathbf{s}(k+n)),$$

by expanding the summation, we have $\forall p, M \in Z^+$ and $\forall k \geq 0$,

$$\begin{aligned}
&\mathbf{x}(k+pM) = \mathbf{A}^{pM}\mathbf{x}(k) + \\
&\mathbf{A}^{pM-1}\mathbf{B}(\mathbf{u} - \mathbf{s}(k)) + \cdots + \mathbf{A}^{pM-1-(M-1)}\mathbf{B}(\mathbf{u} - \mathbf{s}(k-(M-1))) + \\
&\mathbf{A}^{pM-1-M}\mathbf{B}(\mathbf{u} - \mathbf{s}(k-M)) + \cdots + \mathbf{A}^{pM-1-(2M-1)}\mathbf{B}(\mathbf{u} - \mathbf{s}(k-(2M-1))) + \cdots + \\
&\mathbf{A}^{pM-1-(p-1)M}\mathbf{B}(\mathbf{u} - \mathbf{s}(k-(p-1)M)) + \cdots + \mathbf{A}^{pM-1-(pM-1)}\mathbf{B}(\mathbf{u} - \mathbf{s}(k-(pM-1)))
\end{aligned}$$

This further implies that $\forall p, M \in Z^+$ and $\forall k \geq 0$,

$$\begin{aligned}
&\mathbf{x}(k+pM) = \mathbf{A}^{pM}\mathbf{x}(k) + \\
&\mathbf{A}^{(p-1)M}\mathbf{A}^{M-1}\mathbf{B}(\mathbf{u} - \mathbf{s}(k)) + \cdots + \mathbf{A}^{(p-1)M}\mathbf{B}(\mathbf{u} - \mathbf{s}(k-(M-1))) + \\
&\mathbf{A}^{(p-2)M}\mathbf{A}^{M-1}\mathbf{B}(\mathbf{u} - \mathbf{s}(k-M)) + \cdots + \mathbf{A}^{(p-2)M}\mathbf{B}(\mathbf{u} - \mathbf{s}(k-(2M-1))) + \cdots + \\
&\mathbf{A}^{0}\mathbf{A}^{M-1}\mathbf{B}(\mathbf{u} - \mathbf{s}(k-(p-1)M)) + \cdots + \mathbf{A}^{0}\mathbf{B}(\mathbf{u} - \mathbf{s}(k-(pM-1)))
\end{aligned}$$

From (i), since $\mathbf{s}(k_0 + qM + n) = \mathbf{s}(k_0 + n)$ for $q = 0,1,\cdots, p-1$ and for $n = 0,1,\cdots, M-1$, we have $\forall p, M \in Z^+$,

$$\mathbf{x}(k_0 + pM) = \mathbf{A}^{pM}\mathbf{x}(k_0) + \left(\mathbf{A}^{(p-1)M} + \mathbf{A}^{(p-2)M} + \cdots + \mathbf{A}^{0}\right)\sum_{n=0}^{M-1} \mathbf{A}^{M-1-n}\mathbf{B}(\mathbf{u} - \mathbf{s}(k_0 + n)),$$

From equation (3.3), we have $\forall p, M \in Z^+$ and $\forall k \geq 0$,

$$\mathbf{x}(k_0 + pM) = \mathbf{T}\mathbf{D}^{pM}\mathbf{T}^{-1}\mathbf{x}(k_0) + \mathbf{T}\left(\mathbf{D}^{(p-1)M} + \mathbf{D}^{(p-2)M} + \cdots + \mathbf{D}^{0}\right)\sum_{n=0}^{M-1} \mathbf{D}^{M-1-n}\mathbf{T}^{-1}\mathbf{B}(\mathbf{u} - \mathbf{s}(k_0 + n)),$$

which further implies that

$$\mathbf{x}(k_0 + pM) = \mathbf{T}\mathbf{D}^{pM}\mathbf{T}^{-1}\mathbf{x}(k_0) + \sum_{n=0}^{M-1} \mathbf{T}\mathbf{D}^{M-1-n}\left(\sum_{m=0}^{p-1} \mathbf{D}^{mM}\right)\mathbf{T}^{-1}\mathbf{B}(\mathbf{u} - \mathbf{s}(k_0 + n)).$$

Hence, from (3.7), we have:

$$\lim_{p \to +\infty} \mathbf{x}(k_0 + pM) = \mathbf{x}_0^*.$$

From (3.8), we have

$$\mathbf{x}_{i+1}^* = \mathbf{A}\mathbf{x}_i^* + \mathbf{B}(\mathbf{u} - \mathbf{s}(k_0 + i)) \text{ for } i = 0,1,\cdots, M-2.$$

Hence,

$$\lim_{k\to+\infty} \mathbf{x}(kM + k_0 + i) = \mathbf{x}_i^* \text{ for } i = 0,1,\cdots,M-1$$

and completes the proof of this part.

For (ii) implies (i), since

$$\lim_{k\to+\infty} \mathbf{x}(kM + k_0 + i) = \mathbf{x}_i^* \text{ for } i = 0,1,\cdots,M-1,$$

then $\exists k_1 \geq 0$ such that

$$Q(\mathbf{x}(kM + k_0 + i)) = Q(\mathbf{x}_i^*) \text{ for } k \geq k_1 \text{ and } i = 0,1,\cdots,M-1.$$

This implies that $\mathbf{s}(k_0 + kM + i) = \mathbf{s}(k_0 + i)$ for $k \geq k_1$. This completes the proof of this part.

For (ii) implies (iii), since

$$\lim_{k\to+\infty} \mathbf{x}(kM + k_0 + i) = \mathbf{x}_i^* \text{ for } i = 0,1,\cdots,M-1,$$

then $\exists k_1 \geq 0$ such that

$$Q(\mathbf{x}(kM + k_0 + i)) = Q(\mathbf{x}_i^*) \text{ for } k \geq k_1 \text{ and } i = 0,1,\cdots,M-1.$$

By the definition of Ξ_1 in (iii), we have $\mathbf{x}(0) \in \Xi_1$. Hence, this completes the proof of this part.

For (iii) implies (i), since $\mathbf{x}(0) \in \Xi_1$, this implies that

$$Q(\mathbf{x}(kM + k_0 + i)) = Q(\mathbf{x}_i^*) \text{ for } k \geq 0 \text{ and for } i = 0,1,\cdots,M-1$$

is equivalent to $\mathbf{s}(k_0 + kM + i) = \mathbf{s}(k_0 + i)$ for $k \geq 0$. Hence, it completes the proof of this part. And this completes the whole proof of the lemma. ■

APPENDIX B

According to Lemma 1, if a periodic sequence is admissible, then

$$\mathbf{s}(k_0 + M + i) = \mathbf{s}(k_0 + i) \;\; \forall i \geq 0 .$$

Since

$$\begin{aligned}
\mathbf{x}_0^* &= \sum_{n=0}^{M-1} \mathbf{T}\mathbf{D}^{M-1-n} \left(\lim_{p \to +\infty} \sum_{m=0}^{p-1} \mathbf{D}^{mM} \right) \mathbf{T}^{-1}\mathbf{B}(\mathbf{u} - \mathbf{s}(k_0 + n)) \\
&= \sum_{n=0}^{M-1} \mathbf{T} \left(\lim_{p \to +\infty} \sum_{m=0}^{p-1} \mathbf{D}^{mM} \right) \mathbf{D}^{M-1-n}\mathbf{T}^{-1}\mathbf{B}(\mathbf{u} - \mathbf{s}(k_0 + n)) \\
&= \mathbf{T} \left(\lim_{p \to +\infty} \sum_{m=0}^{p-1} \mathbf{D}^{mM} \right) \sum_{n=0}^{M-1} \mathbf{D}^{M-1-n}\mathbf{T}^{-1}\mathbf{B}(\mathbf{u} - \mathbf{s}(k_0 + n)) \\
&= \left(\lim_{p \to +\infty} \sum_{m=0}^{p-1} \mathbf{A}^{mM} \right) \sum_{n=0}^{M-1} \mathbf{A}^{M-1-n}\mathbf{B}(\mathbf{u} - \mathbf{s}(k_0 + n)) \\
&= (\mathbf{I} - \mathbf{A}^M)^{-1} \left(\mathbf{I} - \lim_{p \to +\infty} \mathbf{A}^{pM} \right) \sum_{n=0}^{M-1} \mathbf{A}^{M-1-n}\mathbf{B}(\mathbf{u} - \mathbf{s}(k_0 + n)) \\
&= (\mathbf{I} - \mathbf{A}^M)^{-1} \sum_{n=0}^{M-1} \mathbf{A}^{M-1-n}\mathbf{B}(\mathbf{u} - \mathbf{s}(k_0 + n))
\end{aligned} ,$$

We have

$$\mathbf{x}_0^* = \mathbf{A}^M \mathbf{x}_0^* + \sum_{n=0}^{M-1} \mathbf{A}^{M-1-n}\mathbf{B}(\mathbf{u} - \mathbf{s}(k_0 + n)).$$

As

$$\mathbf{x}_{i+1}^* = \mathbf{A}\mathbf{x}_i^* + \mathbf{B}(\mathbf{u} - \mathbf{s}(k_0 + i)) \text{ for } i = 0,1,\cdots,M-2 ,$$

this implies that

$$\mathbf{x}_i^* = \mathbf{A}^M \mathbf{x}_i^* + \sum_{n=0}^{M-1} \mathbf{A}^{M-1-n}\mathbf{B}(\mathbf{u} - \mathbf{s}(k_0 + i + n)) \text{ for } i = 0,1,\cdots,M-1$$

$$= (\mathbf{I} - \mathbf{A}^M)^{-1} \sum_{n=0}^{M-1} \mathbf{A}^{M-1-n}\mathbf{B}(\mathbf{u} - \mathbf{s}(k_0 + i + n)) \text{ for } i = 0,1,\cdots,M-1$$

$$= (\mathbf{I} - \mathbf{A}^M)^{-1} \sum_{n=0}^{M-1} \mathbf{A}^{M-1-n}\mathbf{B}\mathbf{u} - (\mathbf{I} - \mathbf{A}^M)^{-1} \sum_{n=0}^{M-1} \mathbf{A}^{M-1-n}\mathbf{B}\mathbf{s}(k_0 + i + n) \text{ for } i = 0,1,\cdots,M-1$$

$$= (\mathbf{I} - \mathbf{A}^M)^{-1} (\mathbf{I} - \mathbf{A}^M)(\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\mathbf{u} - (\mathbf{I} - \mathbf{A}^M)^{-1} \sum_{j=0}^{M-1} \mathbf{A}^{\bmod(i-1-j,M)}\mathbf{B}\mathbf{s}(k_0 + j) \text{ for } i = 0,1,\cdots,M-1$$

$$= \left(\mathbf{I}-\mathbf{A}\right)^{-1}\mathbf{Bu} - \sum_{j=0}^{M-1}\mathbf{A}^{\bmod(i-1-j,M)}\left(\mathbf{I}-\mathbf{A}^{M}\right)^{-1}\mathbf{B}\mathbf{s}(k_0+j) \text{ for } i=0,1,\cdots,M-1.$$

This further implies that

$$\begin{bmatrix}\mathbf{x}_0^* \\ \mathbf{x}_1^* \\ \vdots \\ \mathbf{x}_{M-1}^*\end{bmatrix} = \begin{bmatrix}\left(\mathbf{I}-\mathbf{A}\right)^{-1}\mathbf{Bu} \\ \left(\mathbf{I}-\mathbf{A}\right)^{-1}\mathbf{Bu} \\ \vdots \\ \left(\mathbf{I}-\mathbf{A}\right)^{-1}\mathbf{Bu}\end{bmatrix} - \begin{bmatrix}\mathbf{A}^{M-1}\left(\mathbf{I}-\mathbf{A}^{M}\right)^{-1}\mathbf{B} & \mathbf{A}^{M-2}\left(\mathbf{I}-\mathbf{A}^{M}\right)^{-1}\mathbf{B} & \cdots & \left(\mathbf{I}-\mathbf{A}^{M}\right)^{-1}\mathbf{B} \\ \left(\mathbf{I}-\mathbf{A}^{M}\right)^{-1}\mathbf{B} & \mathbf{A}^{M-1}\left(\mathbf{I}-\mathbf{A}^{M}\right)^{-1}\mathbf{B} & \cdots & \mathbf{A}\left(\mathbf{I}-\mathbf{A}^{M}\right)^{-1}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}^{M-2}\left(\mathbf{I}-\mathbf{A}^{M}\right)^{-1}\mathbf{B} & \mathbf{A}^{M-3\left(\mathbf{I}-\mathbf{A}^{M}\right)^{-1}}\mathbf{B} & \cdots & \mathbf{A}^{M-1}\left(\mathbf{I}-\mathbf{A}^{M}\right)^{-1}\mathbf{B}\end{bmatrix}\begin{bmatrix}\mathbf{s}(k_0) \\ \mathbf{s}(k_0+1) \\ \vdots \\ \mathbf{s}(k_0+M-1)\end{bmatrix}$$

and

$$Q\left(\begin{bmatrix}\left(\mathbf{I}-\mathbf{A}\right)^{-1}\mathbf{Bu} \\ \left(\mathbf{I}-\mathbf{A}\right)^{-1}\mathbf{Bu} \\ \vdots \\ \left(\mathbf{I}-\mathbf{A}\right)^{-1}\mathbf{Bu}\end{bmatrix} - \begin{bmatrix}\mathbf{A}^{M-1}\left(\mathbf{I}-\mathbf{A}^{M}\right)^{-1}\mathbf{B} & \mathbf{A}^{M-2}\left(\mathbf{I}-\mathbf{A}^{M}\right)^{-1}\mathbf{B} & \cdots & \left(\mathbf{I}-\mathbf{A}^{M}\right)^{-1}\mathbf{B} \\ \left(\mathbf{I}-\mathbf{A}^{M}\right)^{-1}\mathbf{B} & \mathbf{A}^{M-1}\left(\mathbf{I}-\mathbf{A}^{M}\right)^{-1}\mathbf{B} & \cdots & \mathbf{A}\left(\mathbf{I}-\mathbf{A}^{M}\right)^{-1}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}^{M-2}\left(\mathbf{I}-\mathbf{A}^{M}\right)^{-1}\mathbf{B} & \mathbf{A}^{M-3}\left(\mathbf{I}-\mathbf{A}^{M}\right)^{-1}\mathbf{B} & \cdots & \mathbf{A}^{M-1}\left(\mathbf{I}-\mathbf{A}^{M}\right)^{-1}\mathbf{B}\end{bmatrix}\begin{bmatrix}\mathbf{s}(k_0) \\ \mathbf{s}(k_0+1) \\ \vdots \\ \mathbf{s}(k_0+M-1)\end{bmatrix}\right) = \begin{bmatrix}\mathbf{s}(k_0) \\ \mathbf{s}(k_0+1) \\ \vdots \\ \mathbf{s}(k_0+M-1)\end{bmatrix}.$$

Hence, equation (3.15) is satisfied and this completes the proof. ■

APPENDIX C

Denote $\mathbf{Q} \equiv \mathbf{I} - \mathbf{A}^P$. Since $\mathbf{A} = \mathbf{T}\mathbf{D}\mathbf{T}^{-1}$, we have:

$$\mathbf{Q}\mathbf{T} = \left(\mathbf{I} - \mathbf{A}^P\right)\mathbf{T} = \left(\mathbf{T}\mathbf{T}^{-1} - \mathbf{T}\mathbf{D}^P\mathbf{T}^{-1}\right)\mathbf{T} = \mathbf{T}\left(\mathbf{I} - \mathbf{D}^P\right)\mathbf{T}^{-1}\mathbf{T} = \mathbf{T}\left(\mathbf{I} - \mathbf{D}^P\right).$$

As $\lambda_{i+N-n_d} = e^{\frac{j2\pi k_i}{P}}$ for $k_i \in Z$ and $i = 1,2,\cdots,n_d$, we have

$$\mathbf{Q}\mathbf{T} = \mathbf{T}\begin{bmatrix} 1-\lambda_1^P & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & & \vdots \\ \vdots & \ddots & 1-\lambda_{N-n_d}^P & \ddots & & \vdots \\ \vdots & & \ddots & 0 & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & 0 \end{bmatrix}.$$

Since $\mathbf{T} = \left[\xi_1, \quad \cdots, \quad \xi_N\right]$, we have

$$\mathbf{Q}\mathbf{T} = \left[\left(1-\lambda_1^P\right)\xi_1, \quad \cdots, \quad \left(1-\lambda_{N-n_d}^P\right)\xi_{N-n_d}, \quad \mathbf{0}, \quad \cdots, \quad \mathbf{0}\right]$$

and

$$rank(\mathbf{Q}\mathbf{T}) = rank\left(\left[\left(1-\lambda_1^P\right)\xi_1, \quad \cdots, \quad \left(1-\lambda_{N-n_d}^P\right)\xi_{N-n_d}, \quad \mathbf{0}, \quad \cdots, \quad \mathbf{0}\right]\right).$$

Since $\mathbf{T}$ is a full rank matrix, $\left\{\xi_1,\cdots,\xi_{N-n_d}\right\}$ are linearly independent. As $1-\lambda_i^P \neq 0$ for $i = 1,2,\cdots,N-n_d$, $rank(\mathbf{Q}\mathbf{T}) = N - n_d$. However, $rank(\mathbf{Q}\mathbf{T}) \leq rank(\mathbf{Q})$. Hence, $rank(\mathbf{Q}) \geq N - n_d$. Since

$$\mathbf{Q} = \left[\left(1-\lambda_1^P\right)\xi_1, \quad \cdots, \quad \left(1-\lambda_{N-n_d}^P\right)\xi_{N-n_d}, \quad \mathbf{0}, \quad \cdots, \quad \mathbf{0}\right]\mathbf{T}^{-1},$$

$rank(\mathbf{Q}) \leq N - n_d$. Hence, $rank(\mathbf{Q}) = N - n_d$. As a result, the number of linearly independent rows in the matrix $\mathbf{I} - \mathbf{A}^P$ is $N - n_d$.

Since $\Psi_P \neq \emptyset$, $\exists \mathbf{x}(0) \in \Re^N$ such that $\mathbf{r}_i\mathbf{x}(k_0) = L_i$ for $i = 1,2,\cdots,N-n_d$. As $\sum_{i=1}^{N-n_d} c_{i,n}L_i = L_{N-n_d+n}$ for $n = 1,2,\cdots,n_d$, $\sum_{i=1}^{N-n_d} c_{i,n}\mathbf{r}_i\mathbf{x}(k_0) = L_{N-n_d+n}$ for $n = 1,2,\cdots,n_d$. Since $\sum_{i=1}^{N-n_d} c_{i,n}\mathbf{r}_i = \mathbf{r}_{N-n_d+n}$ for $n = 1,2,\cdots,n_d$, $\mathbf{r}_{N-n_d+n}\mathbf{x}(k_0) = L_{N-n_d+n}$ for $n = 1,2,\cdots,n_d$. Hence, $\mathbf{r}_i\mathbf{x}(k_0) = L_i$ for $n = 1,2,\cdots,N$. This implies that

$$\left(\mathbf{I}-\mathbf{A}^{P}\right)\mathbf{x}(k_0)=\sum_{j=0}^{P-1}\mathbf{A}^{P-1-j}\mathbf{B}\left(\mathbf{u}(k_0+j)-\mathbf{s}(k_0+j)\right),$$

or

$$\mathbf{x}(k_0)=\mathbf{A}^{P}\mathbf{x}(k_0)+\sum_{j=0}^{P-1}\mathbf{A}^{P-1-j}\mathbf{B}\left(\mathbf{u}(k_0+j)-\mathbf{s}(k_0+j)\right).$$

From (5.8), we have

$$\mathbf{x}(k_0+P)=\mathbf{A}^{P}\mathbf{x}(k_0)+\sum_{j=0}^{P-1}\mathbf{A}^{P-1-j}\mathbf{B}\left(\mathbf{u}(k_0+j)-\mathbf{s}(k_0+j)\right),$$

we have $\mathbf{x}(k_0)=\mathbf{x}(k_0+P)$. Hence, the SDMs exhibit limit cycle behaviors with period P for $k\geq k_0$. And Ψ_P is the corresponding nonempty set of initial conditions.

When $\Psi_P=\emptyset$ or $\exists n\in\{1,2,\cdots,n_d\}$ such that $\sum_{i=1}^{N-n_d}c_{i,n}L_i\neq L_{N-n_d+n}$, then there does not exist $\mathbf{x}(0)$ such that $\mathbf{x}(k_0+P)=\mathbf{x}(k_0)$. Hence, there will not exist any fixed point or periodic state sequence, and this completes the proof. ■

APPENDIX D

If $|x'(k)| > |b_N|$, or $|x'(k)| < |b_N|$ and $Q(a_N b_N) = -1$, then the forward and backward dynamics of the SDMs are defined. Consequently, $\wp$ is well defined. $\forall \mathbf{x}(0) \in \wp$, $\Im(\mathbf{x}(0)) \in \wp$ because the forward dynamics exists. This implies $\Im(\wp) \subseteq \wp$. $\forall \mathbf{x}(0) \in \wp$, $\exists \mathbf{x}(-1) \in \wp \quad \Im(\mathbf{x}(-1)) = \mathbf{x}(0)$ because the backward dynamics exists. This implies $\Im(\wp) \supseteq \wp$. Hence, $\Im(\wp) = \wp$ and $\wp$ is an invariant set under the system mapping. This completes the proof. ■

APPENDIX E

It can be seen that $\forall \overline{u} \in \Re$, $\forall \mathbf{x}(0) \in \Re^N$, $\forall a_i \in \Re$ for $i = 0,1,\cdots,N$, $\forall b_j \in \Re$ for $j = 1,\cdots,N$, $\forall k_0 \geq 0$ and $\forall \mathbf{x}^c(k_0+1) \in B_0$, $\mu_{\text{distance}}\left(\mathbf{x}^c(k_0+1)\right) > 0$ and $\mu_{\text{stable}}\left(\mathbf{x}^c(k_0+1)\right) > 0$. If $PER(k_0) \bigcap B_0 = B_0$ or $PER(k_0) \bigcap B_0 = \emptyset$, then $\mu_{\text{aperiodic}}\left(\mathbf{x}^c(k_0+1)\right) > 0$. Although $\mu_{\text{aperiodic}}\left(\mathbf{x}^c(k_0+1)\right) = 0$ if $PER(k_0) \bigcap B_0 \neq B_0$, $PER(k_0) \bigcap B_0 \neq \emptyset$ and $\mathbf{x}(k_0+1) \in B_0 \bigcap PER(k_0)$, since $PER(k_0) \bigcap B_0 \neq B_0$, $\exists \mathbf{x}^c(k_0+1) \in B_0 \setminus PER(k_0)$ such that $\mu_{\text{aperiodic}}\left(\mathbf{x}^c(k_0+1)\right) > 0$. Hence, $\exists \mathbf{x}^c(k_0+1) \in B_0 \setminus PER(k_0)$ such that $\mu_{\mathbf{x}^c(k_0+1)}\left(\mathbf{x}^c(k_0+1)\right) > 0$. As a result, if $\mathbf{A}\mathbf{x}(k_0) + \mathbf{B}(\overline{\mathbf{u}} - Q(\mathbf{x}(k_0))) \in \Re^N \setminus B_0$, then the fuzzy impulsive controller will reset the system state of the loop filter to $\mathbf{x}^c(k_0+1)$ where $\mathbf{x}^c(k_0+1) \in B_0$. If $\mathbf{A}\mathbf{x}(k_0) + \mathbf{B}(\overline{\mathbf{u}} - Q(\mathbf{x}(k_0))) \in B_0$, since no control force is applied to the SDM, $\mathbf{x}^c(k_0+1) = \mathbf{x}(k_0+1) \in B_0$. Hence, $\mathbf{x}^c(k) \in B_0$ for $k > k_0$. Thus, $\mathbf{x}^c(k) \in B_0$ for $k > 0$. And this completes the proof. ■

APPENDIX F

Since $\forall \bar{u} \in \Re$, $\forall \mathbf{x}(0) \in \Re^N$, $\forall a_i \in \Re$ for $i = 0,1,\cdots,N$ and $\forall b_j \in \Re$ for $j = 1,\cdots,N$, $\mathbf{x}^c(k) \in B_0$ for $k > 0$. Since the maximum distance between any points in B_0 is less than or equal to $2V_{cc}\sqrt{N}$, this completes the proof. ■

APPENDIX G

Since $PER(k) \cap B_0 \neq B_0$, $\mu_{\text{aperiodic}}(\mathbf{x}^c(k+1)) \neq 0$. Hence, $\exists M > 0$ such that $\mathbf{x}^c(k) = \mathbf{x}^c(k+M)$ for $k > k_0$, and this completes the proof. ■

REFERENCES

[1] Candy, J. C.. (1974). A use of limit cycle oscillations to obtain robust analog-to-digital converters. *IEEE Transactions on Communications*, COM-22(3):298-305.

[2] Aziz, P. M., Sorensen, H. V. & Spiegel, J. vn der. (1996). An overview of sigma-delta converters. *Signal Processing Magazine*, 13(1):61-84.

[3] Jayaraman, A., Chen, P. F., Hanington, G., Larson, L. & Asbeck, P. (1998). Linear high-efficiency microwave power amplifiers using bandpass delta-sigma modulators. *IEEE Microwave and Guided Wave Letters*, 8(3):121-123.

[4] Janssen, E. & Reefman, D.. (2003). Super-audio CD: an introduction. *IEEE Signal Processing Magazine*, 20(4):83-90.

[5] Kawahito, S., Cerman, A., Aramaki, K. & Tadokoro, Y.. (2003). A weak magnetic field measurement system using micro-fluxgate sensors and delta-sigma interface. *IEEE Transactions on Instrumentation and Measurement*, 52(1):103-110.

[6] Rosa, J. M. de la, Pérez-Verdú, B., Río, R. del & Rodríguez-Vázquez, A.. (2000). A CMOS 0.8-μm transistor-only 1.63-MHz switched-current bandpass ΣΔ modulator for AM signal A/D conversion. *IEEE Journal of Solid-State Circuits*, 35(8):1220-1226.

[7] Gerosa, A., Maniero, A. & Neviani, A.. (2004). A fully integrated two-channel A/D interface for the acquisition of cardiac signals in implantable pacemakers. *IEEE Journal of Solid-State Circuits*, 39(7):1083-1093.

[8] Candy, J. C.. (1985). A use of double integration in sigma delta modulation. *IEEE Transactions on Communications*, 33(3):249-258.

[9] Jager, F. De.. (1952). Delta modulation – a method of PCM transmission using the one unit code. Philips Research, Rep. 7, 442-466.

[10] Steele, R. (1975). *Delta modulation systems*. Pentech Press, London, 320p.

[11] Cutler. C. C. (1960). Transmission systems employing quantization. U.S. Patent no. 2,927,962, 8 Mar. 1960.

[12] Inose, H. & Yasuda, Y.. (1963). A unity bit coding method by negative feedback. *Proceedings of the IEEE*, 1524-1535, Nov. 1963.

[13] Brahm, C. B.. (1965). Feedback integrating system. U.S. Patent no. 3,192,371, 29 Jun. 1965.

[14] Miura, T., Shi, K. & Iwata, J.. (1971). Signal conversion system with storage and correction of quantization error. U.S. Patent 3,560,957, 2 Feb. 1971.

[15] Booton, R. C. (1952). Nonlinear control systems with statistical inputs. *MIT Dynamica Analysis and Control Laboratory*, Report no. 61, Mar. 1952.

[16] Atherton, D. P. (1975). *Nonlinear control engineering*. London:Van Nostrand Reinhold, 627p.

[17] Smith, H. W. (1966). Approximate analysis of randomly excited nonlinear controls. *Research Monograph*, no. 34, MIT Press, Cambridge, Massachusetts.

[18] Ardalan, S. H. & Paulos, J. J. (1987). An analysis of nonlinear behavior in delta-sigma modulators. *IEEE Transactions on Circuits and Systems*, CAS-34(6):593-603.

[19] Mees, A. I. & Bergen, A. R. (1975). Describing functions revisited. *IEEE Transactions on Automatic Control*, AC-20(4):473-478.

[20] Hein, S. & Zakhor, A.. (1993). On the stability of sigma delta modulators. *IEEE Transactions on Signal Processing*, 41(7):2322-2348.

[21] Ho, C. Y. F., Ling, B. W. K. & Reiss, J. D. (2006). Stability of sinusoidal responses of marginally stable bandpass sigma delta modulators. *International Journal of Circuit Theory and Applications*, 34(6):593-605.

[22] Stikvoort, E. F. (1988). Some remarks on the stability and performance of the noise shaper or sigma-delta modulator. *IEEE Transactions on Communications*, 36(10):1157-1162.

[23] Baird, R. T. & Fiez, T. S. (1994). Stability analysis of high-order delta-sigma modulation for ADC's. *IEEE Transactions on Circuits and Systems – II: Analog and Digital Signal Processing*, 41(1):59-61.

[24] Schreier, R., Goodson, M. V. & Zhang, B.. (1997). An algorithm for computing convex positively invariant sets for delta-sigma modulators. *IEEE Transactions on Circuits and Systems – I: Fundamental Theory and Applications*, 44(1):38-44.

[25] Feely, O.. (1997). A tutorial introduction to nonlinear dynamics and chaos and their application to sigma delta modulators. *International Journal of Circuit Theory and Applications*, 25:347-367.

[26] Chua, L. O.. (1988). Chaos in digital filters. *IEEE Transactions on Circuits and Systems*, 35(6):648-1070.

[27] Ling, B. W. K., Tam, P. K. S. & Yu, X.. (2003). Step response of a second-order digital filter with two's complement arithmetic. *IEEE Transactions on Circuits and Systems – I: Regular Papers*, 50(4):510-522.

[28] Oppenheim, A. V., Schafer, R. W. & Buck, J. R.. (1999). *Discrete-time signal processing*. 2nd Ed., Prentice Hall, Upper Saddle River, NJ, London, 860p.

[29] Hauser, M. W.. (1991). Principles of oversampling A/D conversion. *Journal of the Audio Engineering Society*, 39(7/8):3-26.

[30] Friedman, V.. (1988). The structure of the limit cycles in sigma delta modulation. *IEEE Transactions on Communications*, 36(8):972-979.

[31] Sripad, A. B. & Synder, D. L.. (1977). A necessary and sufficient condition for quantization errors to be uniform and white. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 25(5):442-448.

[32] Candy, J. C. & Benjamin, O. J.. (1981). The structure of quantization noise from sigma-delta modulation. *IEEE Transactions on Communications*, COM-29(9):1316-1323.

[33] Feely, O.. (1995). Theory of lowpass and bandpass sigma-delta modulation. *IEE Colloquium on Oversampling and Sigma-Delta Strategies for DSP*, 7/1-7/8, 23 Nov 1995.

[34] Mann, S. I. & Taylor, D. P.. (1999). Limit cycle behavior in the double-loop bandpass Σ-Δ A/D converter. *IEEE Transactions on Circuits and Systems – II: Analog and Digital Signal Processing*, 46(8):1086-1089.

[35] Hein, S.. (1993). Exploiting chaos to suppress spurious tones in general double-loop ΣΔ modulators. *IEEE Transactions on Circuits and Systems – II: Analog and Digital Signal Processing*, 40(10):651-659.

[36] Feely, O. & Chua, L. O.. (1991). The effect of integrator leak in Σ-Δ modulation. *IEEE Transactions on Circuits and Systems*, 38(11):1293-1305.

[37] Schreier, R. (1994). On the use of chaos to reduce idle-channel tones in delta sigma modulators. *IEEE Transactions on Circuits and Systems—I:Fundamental Theory and Applications*, 41(8):539-547.

[38] Zames, G. & Shneydor, N. A. (1976).Dithering in nonlinear systems. *IEEE Transactions on Automatic Control*, AC-21(5):660-667.

[39] Hyun, D. & Fischer, G. (2002). Limit cycles and pattern noise in single-stage single-bit delta-sigma modulators. *IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications*, 49(5):646-656.

[40] Risbo, L. (1995). On the design of tone-free ΣΔ modulators. *IEEE Transactions on Circuits and Systems—II: Analog and Digital Signal Processing*, 42(1):52-55.

[41] Feng, G. (2002). Stability analysis of piecewise discrete-time linear systems. *IEEE Transactions on Automatic Control*, 47(7):1108-1115.

[42] Zourntos, T. & Johns, D. A.. (2002). Variable-structure compensation of delta-sigma modulators: stability and performance. *IEEE Transactions on Circuits and Systems – I: Fundamental Theory and Applications*, 49(1):41-53.

[43] Uçar, A.. (2003). Bounding integrator output of sigma-delta modulator by time delay feedback control. *IEE Proceedings – Circuits, Devices and Systems*, 150(1):31-37.

[44] Wang, L. X. (1997). *A course in fuzzy systems and control*. Upper Saddle River, NJ: Prentice Hall, 424p.

[45] Reefman, D. & Janssen, E.. (2002). Signal processing for direct stream digital: a tutorial for digital sigma delta modulation and 1-bit digital audio processing. Philips Research, Eindhoven, White Paper.

[46] Lampinen, H. & Vainio, O.. (2001). An optimization approach to designing OTAs for low-voltage sigma-delta modulators. *IEEE Transactions on Instrumentation and Measurement*, 50(6):1665-1671.

[47] Márkus, J., Silva, J. & Temes, G. C.. (2004). Theory and applications of incremental ΔΣ converters. *IEEE Transactions on Circuits and Systems – I: Regular Papers*, 51(4):678-690.

[48] Bajdechi, O., Gielen, G. E. & Huijsing, J. H.. (2004). Systematic design exploration of delta-sigma ADCs. *IEEE Transactions on Circuits and Systems – I: Regular Papers*, 51(1):86-95.

[49] Francken, K. & Gielen, G. G. E.. (2003). A high-level simulation and synthesis environment for ΔΣ modulators. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 22(8):1049-1061.

[50] Maulik, P. C., Chadha, M. S., Lee, W. L. & Crawley, P. J.. (2000). A 16-bit 250-kHz delta-sigma demodulator and decimation filter. *IEEE Journal of Solid-State Circuits*, 35(4):458-467.

[51] Abeysekera, S. S., Xue, Y. & Charoensak, C.. (2003). Design of optimal and narrow-band Laguerre filters for sigma-delta demodulators. *IEEE Transactions on Circuits and Systems – II: Analog and Digital Signal Processing*, 50(7):368-375.

[52] Quevedo, D. E. & Goodwin, G. C.. (2005). Multistep optimal analog-to-digital conversion. *IEEE Transactions on Circuits and Systems – I: Regular Papers*, 52(3):503-515.

[53] Marques, A., Peluso, V., Steyaert, M. S. & Sansen, W. M.. (1998). Optimal parameters for ΔΣ modulator topologies. *IEEE Transactions on Circuits and Systems – II: Analog and Digital Signal Processing*, 45(9):1232-1241.

[54] Rombouts, P. & Weyten L.. (2004). Systematic design of double-sampling ΣΔ A/D converters with a modified noise transfer function. *IEEE Transactions on Circuits and Systems – II: Express Briefs*, 51(12):675-679.

[55] Kuo, T.-H., Chen, K.-D. & Chen, J.-R.. (1999). Automatic coefficients design for high-order sigma-delta modulators. *IEEE Transactions on Circuits and Systems – II: Analog and Digital Signal Processing*, 46(1):6-15.

[56] Ho, C. Y.-F., Ling, B. W.-K., Liu, Y.-Q., Tam, P. K.-S. & Teo, K.-L.. (2005). Design of nonuniform near allpass complementary FIR filters via a semi-infinite programming technique. *IEEE Transactions on Signal Processing*, 53(1):376-380.

[57] Ito, S., Liu, Y. & Teo, K. L.. (2000). A dual parametrization method for convex semi-infinite programming. *Annals of Operations Research*, 98:189-213.

[58] Ho, C. Y. F., Ling, B. W. K., Reiss, J. D. & Yu, X.. (2005). Occurrence of Elliptical Fractal Patterns in Multi-bit Bandpass Sigma Delta Modulators. *International Journal of Bifurcation and Chaos*, 15(9):3377-3380.

[59] Ling, B. W. K., Ho, C. Y. F., Reiss, J. D. & Yu, X.. (2005). Nonlinear Behaviors of Bandpass Sigma Delta Modulators with Stable System Matrices. *Proceedings of International Conference of Acoustics, Speech and Signal Processing*, ICASSP, 4:73-76.

[60] Ho, C. Y. F., Ling, B. W. K. & Reiss, J. D.. (2005). Fuzzy Impulsive Control of High Order Sigma Delta Modulators. *Proceedings of the 118th Audio Engineering Society*, 118th AES, Barcelona, 6451.

[61] Ho, C. Y. F., Ling, B. W. K. & Reiss, J. D.. (2005). Fuzzy Impulsive Control of High Order Interpolative Lowpass Sigma Delta Modulators. *IEEE Transactions on Circuits and Systems—I*, 53(10):2224-2233.

[62] Ho, C. Y. F., Ling, B. W. K. & Reiss, J. D.. (2005). Design of Interpolative Sigma Delta Modulators via a Semi-infinite Programming Approach. *Proceedings of the 5th International Conference of Advanced A/D and D/A Conversion Techniques and their Applications*, ADDA, Limerick, 271-276.

[63] Ho, C. Y. F., Ling, B. W. K., Reiss, J. D., Liu, Y. & Teo, K. L.. (2005). Design of Interpolative Sigma Delta Modulators via a Semi-infinite Programming Approach. *IEEE Transactions on Signal Processing*, 54(10):4047-4051.

[64] Steiner, P. & Yang, W.. (1997). A framework for analysis of high-order sigma-delta modulators. *IEEE Transactions on Circuits and Systems – II: Analog and Digital Signal Processing*, 44(1):1-10.

[65] Lipshitz, S. P. & Vanderkooy, J.. (2000). Why professional 1-bit sigma-delta conversion is a bad idea. *Proceedings of The 109th Convention of Audio Engineering Society*, 5188-5198.

[66] Lin, T. & Chua, L. O.. (1991). On chaos of digital filters in the real world. *IEEE Transactions on Circuits and Systems*, 38(5):557-558.

[67] García, J. C. de M. & Armada, A. G. (1999). Effects of bandpass sigma-delta modulation on OFDM signals. *IEEE Transactions on Consumer Electronics*, 45(2):318-326.

[68] Maurino, R. & Mole, P. A. (2000). 200-MHz IF 11-bit fourth-order bandpass ΣΔ ADC in SiGe. *IEEE Journal of Solid-State Circuits*, 35(7):959-967.

[69] Cusinato, P., Stefani, F. & Baschirotto, A.. (2001). Reducing the power consumption in high-speed ΣΔ bandpass modulators. *IEEE Transactions on Circuits and Systems – II: Analog and Digital Signal Processing*, 48(10):952-960.

[70] Cusinato, P., Tonietto, D., Stefani, F. & Baschirotto, A.. (2001). A 3.3-V CMOS 10.7-MHz sixth-order bandpass ΣΔ modulator with 74-dB dynamic range. *IEEE Journal of Solid-State Circuits*, 36(4):629-638.

[71] Gao, W. & Snelgrove, W. M.. (1998). A 950-MHz IF second-order integrated LC bandpass delta-sigma modulator. *IEEE Journal of Solid-State Circuits*, 33(5):723-732.

[72] Reiss, J. D.. (2001). *The Analysis of Chaotic Time Series*. PhD Thesis presented to The Academic Faculty, Georgia Institute of Technology, 219p, Atlanta, 87.

[73] Lu, W. S.. (1999). Design of stable IIR digital filters with equiripple passbands and peak-constrained least squares stopbands. *IEEE Transactions on Circuits and Systems – II: Analog and Digital Signal Processing*, 46(11):1421-1426.

[74] Schreier, R. (2003). *The delta-sigma modulators toolbox version 6.0*, Analog Devices Inc., 1st Jan 2003.

[75] Devaney, R. L. (1989). *An introduction to chaotic dynamical systems*, 2nd Ed, Addison-Wesley Publishing Company, 336p.

[76] Ashwin, P., Fu, X. C. and Deane, J. (2003). Properties of the invariant disk packing in a model bandpass sigma-delta modulator. *International Journal of Bifurcations and Chaos*, 13(3): 631-641.

[77] Chua, L. and Lin, T. (1991) On chaos of digital filters in the real world. *IEEE Transactions on Circuits and Systems*, 38(5):557-558.